A possible intrinsic weakness of AES and other cryptosystems

Anna Rimoldi (rimoldi@science.unitn.it)
Department of Mathematics, University of Trento, Italy.

Massimiliano Sala (maxsalacodes@gmail.com)
Department of Mathematics, University of Trento, Italy.

Ilia Toli (ilia.toli@gmail.com)
Department of Mathematics, Northeastern University, Boston, USA.

Abstract

It has been suggested that the algebraic structure of AES (and other similar block ciphers) could lead to a weakness exploitable in new attacks. In this paper, we use the algebraic structure of AES-like ciphers to construct a novel cipher embedding where the ciphers may lose their non-linearity. We show some examples and we discuss the limitations of our approach.

Keywords: AES, block ciphers, group theory.

Introduction

The Advanced Encryption Standard (AES) [Nat01] is nowadays the most widespread block cipher in commercial applications. It represents the state-of-art in block cipher design and provides an unparalleled level of assurance against all known cryptanalytic techniques, except for its round-reduced versions. It is true that AES (and other modern block ciphers) presents a highly algebraic structure, leading researchers to exploit it for new algebraic attacks, but these tries have been unsuccessful as yet (except for academic reduced versions).

The best that one can hope for a cryptosystem is that all its encryption functions behave in unpredictable way (close to random), in particular we would like that it behaves in a way totally different from linear or affine maps.

A sign of strength for AES is that nobody has been able to show that its encryption functions are any closer to linear maps than arbitrary random functions.

However, it might be possible to extend AES to act on bigger spaces, in such a way that the non-random behavior of AES becomes easier to spot. For example, it was hoped that embedding AES into BES would allow easier [1] polynomial systems to break the ciphers (see [MR02], [TZ05]). Generally speaking, the worst scenario consists of a space *large enough* to make AES linear but *small enough* to allow practical computations. This is probably not possible. Our goal is to find a space *small enough* to allow practical computations but *large enough* to identify a specific behavior of AES, showing that it is closer to linear maps than expected.

In Section 1, after some basic algebraic background, we explain our point of view on *block ciphers*. In particular, we introduce the class of *translation based* cryptosystems, which are ciphers enjoying some interesting algebraic properties. We also briefly describe the three main translation-based cryptosystems: AES, SERPENT and PRESENT.

For completeness, in Section 2 we list the best-known attacks on round-reduced versions of AES.

In Section 3 we provide formal techniques to construct a larger space on which the block cipher can act. We call these techniques *space embeddings*. In the case of translation-based ciphers, these embeddings are designed to lower the non-linearity of the encryption functions. We present one specific embedding and we obtain several results on the rank distributions for matrices in the larger space, which are useful to mount attacks.

In Section 4 we present a larger embedding, that apparently works well with AES and other translation-based systems. The effectiveness of this embedding depends heavily on properties of the mixing-layer.

In Section 5 we outline our approach to attack translation-based ciphers (including AES) with our embeddings. Although we have not been able to find an attack giving satisfactory statistical evidence, we have some partial data suggesting that our methods may work, as reported in [RSB10]

In Section 6 we discuss further on our non-linearity notion:

- first, we report results from [Mai09],[MRS10] on embeddings where the decrease in non-linearity can be formally proved;
- then, we propose alternative embeddings highlighting their flaws;
- finally, with group theory proofs we also show that it is very unlikely that a representation/embedding can completely linearize any version of AES.

[1] *easier* than systems coming from random maps.

1 Preliminaries

In this section we recall well-known results in group theory and finite field theory [LN97] in order to fix the notation we will use in the sequel. We also outline some basic ideas about *block ciphers* and we recall the structure of three well-known cryptosystems: AES, SERPENT and PRESENT.

1.1 Group representations

Let $n \geq 2$ be an integer. Let $V = (\mathbb{F}_2)^n$ be the vector space over the finite field $\mathbb{F}_2$ of dimension n. We denote by $\mathrm{Sym}(V)$ and $\mathrm{Alt}(V)$, respectively, the symmetric and alternating group on V. For any N, we denote by Sym_N and Alt_N, respectively, the symmetric and alternating group on $\{1, \ldots, N\}$. Clearly $\mathrm{Sym}(V)$ is isomorphic to Sym_{2^n} (the same for the alternating group). We denote by $\mathrm{GL}(V)$ the group of all linear permutations of V. We recall the well-known formulas:

$$|\mathrm{Sym}(V)| = 2^n!, \quad |\mathrm{Alt}(V)| = \frac{2^n!}{2} \quad |\mathrm{GL}(V)| = \prod_{h=0}^{n-1}(2^n - 2^h) < 2^{n^2}\,.$$

Given a finite group G, we say that G can be linearized if there is an injective morphism $\rho : G \to \mathrm{GL}(V)$ (this is called a "faithful representation" in representation theory). If G can be linearized, then, for any element $g \in G$, we can compute a matrix M_g corresponding to the action of g over V (via ρ). The matrix computation is easy, since it is enough to evaluate g on a basis of V.
If $\rho : G \to \mathrm{GL}(V)$ is a representation of G on V, then we often write vg instead of $v\rho(g)$, if no confusion arises. Also, G is said to *act linearly* on V, and V is called a *G-module*. The *degree* of the representation is by definition the dimension of V. If we consider Sym_N, we can always linearize Sym_N over V via the so-called *regular* representation as follows.
Let V be a vector space with basis $\{e_1, \ldots, e_N\}$. The regular representation $\rho : \mathrm{Sym}_N \to \mathrm{GL}(V)$ is defined by $(e_i)\rho(g) = e_{ig}$. In other words, any permutation in Sym_N is associated to a permutation $(N \times N)$ matrix (and viceversa). Since any finite group G can be embedded in Sym_N for a smallest N, we can always linearize G using the regular representation. But of course this is huge and usually impractical.

1.2 Finite Fields

For any prime p and any positive $m \in \mathbb{N}$, $\mathbb{F}_{p^m}$ is the field with p^m elements (unique up to field isomorphism). It contains an isomorphic copy of $\mathbb{F}_p$ and can thus be thought as an extension of $\mathbb{F}_p$. On the other hand, we can construct any $\mathbb{F}_{q^s}$ from $\mathbb{F}_q$ with $q = p^m$ elements, as follows.

Let $f \in \mathbb{F}_q[x]$ be an irreducible polynomial of degree m. We can consider the quotient $R = \mathbb{F}_q[x]/(f)$, where (f) is the ideal generated by f in $\mathbb{F}_q[x]$. By considering the natural projection $\pi : \mathbb{F}_q[x] \to R$, we call $\alpha = \pi(x)$ and clearly any element of R can be uniquely expressed as a polynomial in α of degree less than m:

$$R = \left\{ \sum_{i=0}^{m-1} a_i \alpha^i \mid a_i \in \mathbb{F}_q \right\}$$

with the condition $f(\alpha) = 0$.

Theorem 1.1. *$R = \mathbb{F}_q[x]/(f)$ is a field and $R \cong \mathbb{F}_{q^m}$.*

We denote by $\mathbb{F}_q^*$ the multiplicative group of non-zero elements of $\mathbb{F}_q$.

Theorem 1.2. *For any finite field $\mathbb{F}_q$, the multiplicative group $\mathbb{F}_q^*$ is cyclic.*

A generator of the cyclic group $\mathbb{F}_q^*$ is called a *primitive element* of $\mathbb{F}_q$.

Definition 1.3. *An irreducible polynomial $f \in \mathbb{F}_q[x]$ is* primitive *if its roots are primitive elements.*

Note that for any q and m there are indeed irreducible polynomials of degree m over $\mathbb{F}_q$ and some of them are primitive.

1.3 Permutation polynomials

Definition 1.4. *A polynomial $f \in \mathbb{F}_q[x]$ is a* permutation polynomial *of $\mathbb{F}_q$ if the associated polynomial function $f : c \mapsto f(c)$ from $\mathbb{F}_q$ into $\mathbb{F}_q$ is a permutation of $\mathbb{F}_q$. If f is an affine map $f : x \mapsto ax + b$ $(a \neq 0)$, we say that f is a* linear polynomial.

We note the following easy results:

(1) Every linear polynomial over $\mathbb{F}_q$ is a permutation polynomial of $\mathbb{F}_q$.

(2) The monomial x^n is a permutation polynomial of $\mathbb{F}_q$ if and only if

$$\gcd(n, q-1) = 1.$$

Permutation polynomials of $\mathbb{F}_q$ of degree less then q can be combined by the operation of composition and subsequent reduction modulo $x^q - x$. The set of permutation polynomials of $\mathbb{F}_q$ of degree less then q forms a group, which is isomorphic to $\mathrm{Sym}(\mathbb{F}_q)$. Then, the symmetric group $\mathrm{Sym}(\mathbb{F}_q)$ and its subgroups can be represented as groups of permutation polynomials.

Theorem 1.5. *For $q > 2$, the symmetric group $\mathrm{Sym}(\mathbb{F}_q)$ is generated by x^{q-2} and all linear polynomials over $\mathbb{F}_q$.*

1.4 Block ciphers

Block ciphers form an important class of cryptosystems in symmetric key cryptography. These are algorithms that encrypt and decrypt blocks of data (with fixed length [2]) according to a shared secret key. We can formally describe such a cryptosystem using the following definition:

Definition 1.6. *A cryptosystem is a pair $(\mathcal{M}, \mathcal{K})$, where:*

- *$\mathcal{M}$ is a finite set of possible messages (plaintexts, ciphertexts);*
- *$\mathcal{K}$, the key-space, is a finite set of possible keys;*
- *we have encryption and decryption functions for any key $k \in \mathcal{K}$:*

$$\phi_k : \mathcal{M} \to \mathcal{M}, \quad \psi_k : \mathcal{M} \to \mathcal{M}, \quad \phi_k, \psi_k \in \mathrm{Sym}(\mathcal{M})$$

such that

$$\psi_k = (\phi_k)^{-1}.$$

Following the most used structure in modern ciphers, in the previous definition we set that the plaintext space coincides with the ciphertext space. W.l.o.g, we can consider $\mathcal{M} = (\mathbb{F}_q)^r$ and $\mathcal{K} = (\mathbb{F}_q)^\ell$, with r and ℓ positive integers, and we change slightly our previous definition.

Definition 1.7. *Let r and ℓ be natural numbers. Let ϕ be any function*

$$\phi : (\mathbb{F}_q)^r \times (\mathbb{F}_q)^\ell \to (\mathbb{F}_q)^r.$$

For any $k \in (\mathbb{F}_q)^\ell$, we denote by ϕ_k the function

$$\phi_k : (\mathbb{F}_q)^r \to (\mathbb{F}_q)^r, \quad \phi_k(x) = \phi(x, k).$$

We say that ϕ is a **algebraic** block cipher *if ϕ_k is a permutation of $(\mathbb{F}_q)^r$ for any key $k \in (\mathbb{F}_q)^\ell$.*

Under this conditions, we can also consider a block cipher as an indexed set of permutations $(\mathbb{F}_q)^\ell \to \mathrm{Sym}((\mathbb{F}_q)^r)$. Any key $k \in \mathcal{K}$ induces a permutation ϕ_k on $\mathcal{M}$. Since $\mathcal{M}$ is usually $V = (\mathbb{F}_2)^r$ for some $r \in \mathbb{N}$, we can consider $\phi_k \in \mathrm{Sym}(V)$.

To achieve the desired security, most modern block ciphers are iterated ciphers that typically incorporate sequences of permutation and substitution operations. In fact, according to the ideas that Shannon proposed in his seminal paper [Sha49], the encryption process takes as input a plaintext and a

[2] Actually, there is a recent approach that allows a slight change of the block length [CYK09]

random key and so proceeds through N similar rounds. In each round (except possibly for a couple, which may be slightly different) the iterated ciphers perform a non-linear substitution operation (or S-box) on disjoint parts of the input that provides "confusion", followed by a permutation (usually a linear/affine transformation) on the whole data that provides "diffusion". A cryptosystem reaches "confusion" if the relationship between plaintext, ciphertext and key is very complicated. The "diffusion" idea consists of spreading the influence of all parts of the input (plaintext and key) to all parts of the ciphertext. The operations performed in a round form the round function. The round function at the ρ-th round ($1 \leq \rho \leq N$) takes as inputs both the output of the $(\rho-1)$-th round and the subkey $k^{(\rho)}$ (also called round-key). Any round key $k^{(\rho)}$ is constructed starting from a master key[3] k of some specified length, e.g. $k \in \mathcal{K} = (\mathbb{F}_2)^\ell$ (nowadays we have $2^{64} \leq |\mathcal{K}| \leq 2^{256}$). The key schedule is a public algorithm (strictly dependent on the cipher) which constructs $N+1$ subkeys $(k^{(0)}, \ldots, k^{(N)})$.

Several independent formal definitions have been proposed for iterated block ciphers (or subclasses of them). Stinson in [Sti95] gives the following definition of *substitution permutation network* (SPN for short). In [DR02] we can find another class of iterated block cipher, called the *key-alternating* block ciphers.

Now, we consider a more recent definition [CDS09] that defines a class (see Definition 1.9), large enough to include some common ciphers, yet restricted enough to have simple criteria guaranteeing an interesting property of the cipher (for details see Subsection 6.3).

Let $V = (\mathbb{F}_2)^r$ with $r = mb$, $b \geq 2$. The vector space V is a direct sum

$$V = V_1 \oplus \cdots \oplus V_b,$$

where each V_i has the same dimension m (over $\mathbb{F}_2$). For any $v \in V$, we will write $v = v_1 \oplus \cdots \oplus v_b$, where $v_i \in V_i$. Also, we consider the projections $\pi_i : V \to V_i$ mapping $v \mapsto v_i$.

Any $\gamma \in \mathrm{Sym}(V)$ that acts as $v\gamma = v_1\gamma_1 \oplus \cdots \oplus v_b\gamma_b$, for some $\gamma_i \in \mathrm{Sym}(V_i)$, is a bricklayer transformation (a "parallel map") and any γ_i is a brick. The maps γ_i's are traditionally called S-boxes and map γ is called a "parallel S-box". A linear (or affine) map $\lambda : V \to V$ is traditionally called a "Mixing Layer" when used in composition with parallel maps. We denote by σ_v a translation over V.

Definition 1.8. *A linear map $\lambda \in \mathrm{GL}(V)$ is a* proper mixing layer *if no sum of some of the V_i (except $\{0\}$ and V) is invariant under λ.*

We can characterize the "translation based" class by the following

[3] also called session key.

Definition 1.9. *We say that $\mathcal{C}$ is* translation based (tb) *if:*

- *it is the composition of a finite number of rounds, such that any round τ_k can be written*[4] *as $\gamma\lambda\sigma_{\bar{k}}$, where*
  - *γ is a round-dependent bricklayer transformation (but it does not depend on k),*
  - *λ is a round-dependent linear map (but it does not depend on k),*
  - *$\bar{k}$ is in V and depends on both k and the round ($\bar{k}$ is called a "round key");*
- *for at least one round we have (at the same time) that λ is proper and that the map $\mathcal{K} \to V$, $k \mapsto \bar{k}$, is surjective (a "proper" round).*

In [CDS09] the authors gave several non-trivial remarks that can be useful. Let us recall the principal ones.

Remark 1.10. A generalization is obtained by allowing a key-independent permutation at the beginning and/or another at the end. This is the case for example for the SERPENT cipher. Since these permutations have no influence on the cryptanalysis of a cipher, they can be ignored.

Remark 1.11. A round consisting of only a translation is still acceptable, by assuming $\gamma = \lambda = 1_V$ (the identity map on V), although obviously it is not proper. Indeed, we can always assume that the first round is of this kind, otherwise we can remove its γ and λ (Remark 1.10). Then, we can also assume that $0\gamma = 0$, since we can add 0γ to the round key of the previous round.
If the previous round is proper, it remains proper since $\sigma_{0\gamma}$ is a permutation over V.

Remark 1.12. To allow affine mixing layers, rather than linear mixing layers, seems a generalization. However, this case is indeed already present in Definition 1.9, since it is enough to change σ_v to incorporate the "translation part" of the mixing layer.

Remark 1.13. A generalization can be obtained by only requiring *at least one* of the rounds to be of the prescribed form (with a proper mixing layer). Although the authors' results still hold in this more general case, we do not know any interesting cipher of this kind.

Note that some famous ciphers, such as the DES, KASUMI and IDEA ciphers, cannot be seen easily as tb ciphers. Some of them (e.g. DES and KASUMI) are of *Feistel* type. They modify only one half of the cipher state in each round. It has been suggested that the Feistel ciphers suffer from a slow speed of diffusion compared to SPN (or *key-iterated*) ciphers.

In the Subsections 1.5, 1.6, 1.7 we are going to describe respectively AES, SERPENT and PRESENT as translation based cryptosystems[5].

[4] we drop round indexes.

[5] The reader can find a full description of these cryptosystems respectively in [DR02], [ABK98] and [AKL+07]

1.5 The AES-128 cryptosystem

Let $\mathcal{M} = \mathcal{K} = V = (\mathbb{F}_2)^r$ with $r = 128$ and let $x \in \mathcal{M}$ be our plaintext, $k \in \mathcal{K}$ our random key and $y = \phi_k(x)$ the corresponding ciphertext. Before describing the individual components γ, λ and σ_k of the *round function*, we recall (see Section 1.2) that it is possible to identify $(\mathbb{F}_2)^8$ with the field $\mathbb{F}_{2^8}$, via the quotient map $\mathbb{F}_{2^8} \leftrightarrow \mathbb{F}_2[x]/\langle \mathsf{m} \rangle$, where $\mathsf{m} \in \mathbb{F}_2[x]$ is an irreducible polynomial such that $\deg(\mathsf{m}) = 8$. The irreducible (but not primitive) AES polynomial is $\mathsf{m} = x^8 + x^4 + x^3 + x + 1$.

Internally, the AES algorithm's operations are performed on a two-dimensional array of bytes, called the State. It consists of 4 rows and 4 columns and each element of this matrix is one byte (i.e. an element of $\mathbb{F}_{2^8} = \mathbb{F}_{256}$). At the start of the encryption process, the input x (the plaintext) is a vector in V and it is first changed into a 16-byte vector:

$$\nu : (\mathbb{F}_2)^{128} \to (\mathbb{F}_{256})^{16}, \quad \mathsf{x} \mapsto \mathsf{y}.$$

Each round performs its operations on the State and after the last round the State is "unwrapped" and "fills up" the output vector.

A preliminary translation $\sigma_{k^{(0)}}$, where $k^{(0)} \in (\mathbb{F}_2)^r$ is the first round key, is applied to the plaintext to form the input to the (`Round 1`). It means that we can consider a preliminary round (`Round 0`) such that $\gamma = 1_V$ and $\lambda = 1_V$ (see Remark 1.11).
In order to obtain the ciphertext, other $N = 10$ rounds follow.
Let $1 \leq \rho \leq N - 1$. A typical round (`Round` ρ) can be written as the composition [6] $\gamma\lambda\sigma_{k^{(\rho)}}$, where

- the parallel map γ is called SubBytes and it works in parallel to each of the 16 bytes of the data;
- the affine map λ is the composition of two operations known as ShiftRows and MixColumns;
- $\sigma_{k^{(\rho)}}$ is the translation with the session key $k^{(\rho)}$ (this operation is called AddRoundKey).

The last round (`Round` N) is atypical and is characterized by $\gamma\bar{\lambda}\sigma_{k^{(N)}}$ where the affine map $\bar{\lambda}$ is only made by the ShiftRows operation. So we obtain our ciphertext $\mathsf{y} = \phi_k(\mathsf{x})$.
In the following, we analyze the structure of each component of the round function.

[6] Note that the order of the operation is exactly: γ, λ, and then σ_k.

1.5.1 SubBytes

The vector space V is the direct sum $V = V_1 \oplus \cdots \oplus V_{16}$ where each $V_i = (\mathbb{F}_2)^8$ $(1 \leq i \leq 16)$. Any parallel map $\gamma \in \mathrm{Sym}(V)$ acts on an element $v \in V$ as $v\gamma = v_1\gamma_1 \oplus \ldots \oplus v_{16}\gamma_{16}$, where $v_i \in V_i$ and $\gamma_i \in \mathrm{Sym}(V_i)$. The SubBytes operation γ is composed by two transformations: the inversion in $\mathbb{F}_{2^8}$ and an affine transformation.
The *inversion operation* is the *patched inversion*[7] in $\mathbb{F}_{2^8}$ (i.e. $\varphi(x) = x^{254}$).
The *affine transformation* over $\mathbb{F}_2$ consists of an affine mapping $\xi : (\mathbb{F}_2)^8 \to (\mathbb{F}_2)^8$, specified by an 8×8 circulant matrix over $\mathbb{F}_2$ and a translation. The result of inversion is regarded as a vector in $(\mathbb{F}_2)^8$ and the output is given by $y = \xi(x)$, where

$$\begin{pmatrix} y_7 \\ y_6 \\ y_5 \\ y_4 \\ y_3 \\ y_2 \\ y_1 \\ y_0 \end{pmatrix} = \begin{pmatrix} 1\,0\,0\,0\,1\,1\,1\,1 \\ 1\,1\,0\,0\,0\,1\,1\,1 \\ 1\,1\,1\,0\,0\,0\,1\,1 \\ 1\,1\,1\,1\,0\,0\,0\,1 \\ 1\,1\,1\,1\,1\,0\,0\,0 \\ 0\,1\,1\,1\,1\,1\,0\,0 \\ 0\,0\,1\,1\,1\,1\,1\,0 \\ 0\,0\,0\,1\,1\,1\,1\,1 \end{pmatrix} \begin{pmatrix} x_7 \\ x_6 \\ x_5 \\ x_4 \\ x_3 \\ x_2 \\ x_1 \\ x_0 \end{pmatrix} + \begin{pmatrix} 0 \\ 1 \\ 1 \\ 0 \\ 0 \\ 0 \\ 1 \\ 1 \end{pmatrix}$$

1.5.2 Mixing Layer

The map $\lambda : V \to V$ is a composition of two linear operations: ShiftRows and MixColumns. The ShiftRows operation is performed as follows. Any byte (an element of $\mathbb{F}_{2^8}$) in row i of the State, where $0 \leq i \leq 3$, is cyclically shifted (towards left) by i positions, as follows:

$$\begin{array}{|c|c|c|c|} \hline s_0 & s_4 & s_8 & s_{12} \\ \hline s_1 & s_5 & s_9 & s_{13} \\ \hline s_2 & s_6 & s_{10} & s_{14} \\ \hline s_3 & s_7 & s_{11} & s_{15} \\ \hline \end{array} \quad \to \quad \text{ShiftRows} \quad \to \quad \begin{array}{|c|c|c|c|} \hline s_0 & s_4 & s_8 & s_{12} \\ \hline s_5 & s_9 & s_{13} & s_1 \\ \hline s_{10} & s_{14} & s_2 & s_6 \\ \hline s_{15} & s_3 & s_7 & s_{11} \\ \hline \end{array}$$

[7] Since the AES consists of 10 rounds and each round requires 16 S-box computations, the probability of there being no 0-inversions during an encryption is $(255/256)^{160} \approx 0.53$.

In other words, we can describe the ShiftRows operation by the map

$$\mathsf{sh} : (\mathbb{F}_{2^8})^{16} \to (\mathbb{F}_{2^8})^{16}$$

$$(s_0, s_1, \cdots, s_{15}) \mapsto (s_0, s_5, s_{10}, s_{15}, s_4, s_9, s_{14}, s_3, s_8, s_{13}, s_2, s_7, s_{12}, s_1, s_6, s_{11}).$$

We can also represent the ShiftRows operation with the following 16×16 block diagonal matrix

$$S = \begin{pmatrix} I & 0 & 0 & 0 \\ 0 & R & 0 & 0 \\ 0 & 0 & R^2 & 0 \\ 0 & 0 & 0 & R^3 \end{pmatrix} \qquad R = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \end{pmatrix}$$

where the matrix R is a permutation matrix over $\mathbb{F}_{2^8}$ that represents the shift of one row by one position.

In order to describe the MixColumns operation, each column of the State can be treated as a four-term polynomial in $\mathbb{F}_{256}[z]$. Let $c(z)$ be one such polynomial. Then each column is replaced by the result of the multiplication in $\mathbb{F}_{256}[z]/(z^4+1)$ by $a(z)$, $c \mapsto c \cdot a \mod (z^4+1)$,

$$(c_1, c_2, c_3, c_4) \longrightarrow (c_1 \cdot a, c_2 \cdot a, c_3 \cdot a, c_3 \cdot a)\,.$$

Note that $a(z)$ is invertible in $\mathbb{F}_{256}[z]/(z^4+1)$. On the other hand, we can see the MixColumns operation as a 4-block diagonal matrix, each block the same MDS matrix (i.e. all minors are non-zero):

$$\begin{pmatrix} z & z+1 & 1 & 1 \\ 1 & z & z+1 & 1 \\ 1 & 1 & z & z+1 \\ z+1 & 1 & 1 & z \end{pmatrix}$$

Remark 1.14. This MDS property is used to ensure that the number of active S-boxes involved in a differential or linear attack increases rapidly, and the security of the AES against these particular attacks can be established.

Obviously, we can also see the whole Mixing Layer (λ linear operation) as a matrix $\mathbf{M}$. We observe that the order of this matrix is quite small, i.e. $\mathbf{M}^8 = 1$. (Also, both the order of ShiftRows and MixColumns are equal to 4.)

1.6 The SERPENT cryptosystem

Let $\mathcal{M} = V = (\mathbb{F}_2)^r$, with $r = 128$. We consider $\mathcal{K} = (\mathbb{F}_2)^\ell$, with the fixed length $\ell = 128$, although the key is designed with variable length.
The encryption ϕ proceeds by $N = 32$ similar rounds and it works as follows:

- a preliminary permutation is applied $\pi : V \to V$ (this is not used for security, rather to ease the implementation);
- there is a preliminary translation with the first round key;
- $N - 1$ rounds with the same structure are applied, but using a different permutation, each composed of a key translation σ_k, a parallel S-box γ and a linear mixing-layer λ (we denote the round ρ by `Round` ρ, with $\rho = 1, ..., 31$);
- the last round (`Round` 32) follows and it consists of the composition $\gamma\lambda\sigma_k$ where $\lambda = 1_V$;
- a final permutation $\pi^{-1} : V \to V$ is performed.

The decryption process is easily obtained by inverting every step of the encryption, using the inverse of the S-boxes, the inverse of the mixing-layer and the reverse order of the round keys.

Let ρ be a natural number such that $1 \leq \rho \leq 31$. In order to describe a typical round (`Round` ρ) we have to specify how the components γ, λ and σ_k are applied. We note that, after the permutation $\pi : V \to V$, we perform a preliminary translation $\sigma_{k^{(0)}}$, where $k^{(0)} \in (\mathbb{F}_2)^r$ is the first round key.

Let $V = V_1 \oplus \cdots \oplus V_{32}$, where , for any $1 \leq j \leq 32$, each $V_j = (\mathbb{F}_2)^4$. Any $\gamma \in \mathrm{Sym}(V)$ acts as $v\gamma = v_1\gamma_1 \oplus \ldots \oplus v_{32}\gamma_{32}$, where $v_j \in V_j$ and $\gamma_j \in \mathrm{Sym}(V_j)$.

We have to characterize each γ_j (i.e. we have to construct each S-box). The eight S-boxes $S_1, \ldots, S_8$ of SERPENT were built "ad hoc" starting from the 8 fixed S-boxes of DES (see [ABK98]). To each v_j we apply the same S-box $S_{i \mod 8}$, so that $S_{i \mod 8}(v_j)$ lies in $(\mathbb{F}_2)^4$. That is, $\gamma_1 = \gamma_2 = \cdots = \gamma_{32} = S_{i \mod 8}$.

Then the linear transformation λ (described in [ABK98]) and a final translation $\sigma_{k^{(\rho)}}$ are applied. The last round (`Round` 32) is only slightly different. The only difference with a typical round is the replacing of the linear transformation λ by 1_V.

1.7 PRESENT: an ultra-lightweight block cipher

PRESENT is an iterated block cipher that consists of $N = 31$ rounds. Let $\mathcal{M} = V = (\mathbb{F}_2)^r$ with $r = 64$. Let $\mathcal{K} = (\mathbb{F}_2)^\ell$, where ℓ may be equal to 80 or 128. We consider only the PRESENT's version such that $\mathcal{K} = (\mathbb{F}_2)^{80}$, since its authors recommend it in order to have a good performance.
We are going to describe how the *round function* $\gamma\lambda\sigma_{k^{(\rho)}}$ (in the ρ-th typical round) is performed.
As in the AES and SERPENT cryptosystems, the encryption process starts with a preliminary round (`Round 0`) that consists of a parallel map $\gamma = 1_V$, a linear transformation $\lambda = 1_V$ and the translation $\sigma_{k^{(0)}}$, where $k^{(0)} \in (\mathbb{F}_2)^r$ is the first round key. A typical round consists of the non-linear operation, called sBoxLayer, the linear transformation, known as pLayer and the sum with the round key.

The parallel map $\gamma \in \mathrm{Sym}(V)$ used in PRESENT acts as

$$v\gamma = v_1\gamma_1 \oplus \ldots \oplus v_{16}\gamma_{16},$$

where each $v_i \in (\mathbb{F}_2)^4$ and $\gamma_i \in \mathrm{Sym}((\mathbb{F}_2)^4)$ $(1 \leq i \leq 16)$. The action of any brick $\gamma_i : (\mathbb{F}_2)^4 \to (\mathbb{F}_2)^4$ is given by the following table, using an hexadecimal notation:

x	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
$\gamma[x]$	C	5	6	B	9	0	A	D	3	E	F	8	4	7	1	2

The affine map $\lambda : V \to V$ is a bit permutation as given by the following table, where the bit i of the intermediate state is moved to the bit position $P(i)$.

i	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
$P(i)$	0	16	32	48	1	17	33	49	2	18	34	50	3	19	35	51
i	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31
$P(i)$	4	20	36	52	5	21	37	53	6	22	38	54	7	23	39	55
i	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47
$P(i)$	8	24	40	56	9	25	41	57	10	26	42	58	11	27	43	59
i	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63
$P(i)$	12	28	44	60	13	29	45	61	14	30	46	62	15	31	47	63

2 Known attacks

AES's structure has been used to carry out some innovative analysis. Such attacks tend to have a similar form:

- they identify a property holding for a few rounds with a good probability;
- they use special techniques to extend the attack to more rounds.

The following table summarizes the more successful attacks on round-reduced versions of the AES cryptosystem:

Key	Rounds	Texts	Time	Type	Reference
128	5	2^{11}	2^{40}	Square attack	[DR98]
128	5	$2^{29.5}$	2^{31}	Impossible diff.	[BK00]
128	5	2^{39}	2^{39}	Boomerang attack	[Bir04]
128	6	2^{32}	2^{72}	Square attack	[DR98]
128	6	$2^{34.6}$	2^{44}	Partial Sum	[FKL$^+$00]
128	6	$2^{91.5}$	2^{122}	Impossible diff.	[CKK$^+$01]
128	6	2^{71}	2^{71}	Boomerang attack	[Bir04]
128	7	$2^{128} - 2^{119}$	2^{120}	Partial Sum	[FKL$^+$00]
128	7	2^{32}	2^{128}	Collision	[GM00]
192	7	$2^{91.2}$	$2^{139.2}$	Impossible diff.	[**?**]
192	8	2^{127}	2^{188}	Partial Sum	[FKL$^+$00]
192	10	2^{124}	2^{183}	(Related-key) Rectangle	[BDK05]
192	12	2^{123}	2^{176}	(Related-key) Ampl. Boomerang	[BK09]
256	8	2^{32}	2^{194}	Partial Sum	[FKL$^+$00]
256	9	2^{85}	2^{126}	Partial Sum	[FKL$^+$00]
256	10	2^{114}	2^{173}	(Related-key) Rectangle	[BDK05]
256	14	2^{119}	2^{119}	(Related-key) Boomerang	[BK09]

Other researchers attack small scale variants of the AES, where also the message space and the key space are reduced (see e.g.[CW09]). A recent practical attack (due to A.Biryukov, O.Dunkelman, N.Keller, D.Khovratovich, A.Shamir) on a (10-round version) of AES-256 has been presented ([BDK$^+$10]).

3 First results

In the literature there are some ways of representing the same cipher (e.g. AES), like BES [MR02] or Dual Ciphers [BB02], that could be useful for the cryptanalysis. Other ways of representing AES that exploit its structure can be found, for example, in [CMR07].
In this section we represent "AES-like" ciphers by embedding them into larger ciphers. In Subsection 3.1 we begin with We want to enlarge Ω to a set W such that:

(1) W is endowed with a vector space structure;

(2) the permutations can be extended to act linearly on the whole W.

In Subsection 3.2 we provide one specific embedding of AES-like ciphers that linearizes the non-linear part of these ciphers, but it fails to linearize the whole cipher. In particular our embedding can be applied to AES, PRESENT and SERPENT.

3.1 Some preliminary results

Let Ω be a set such that $|\Omega| = n$, let $\mathrm{Sym}(\Omega)$ be the symmetric group on Ω and let W be a vector space over a field $\mathbb{F}$ (not necessarily a finite field).

Definition 3.1. *Let $G \leq \mathrm{Sym}(\Omega)$. An injective map $\phi : \Omega \to W$ is a* **space embedding** *with respect to the group G if, $\forall \sigma \in G$, $\exists A_\sigma \in \mathrm{GL}(W)$ such that $\phi \circ \sigma = A_\sigma \circ \phi$.*

Moreover, $\phi(\Omega)$ is the set of all admissible vectors (w.r.t. ϕ), the subspace $\langle \phi(\Omega) \rangle$ is the admissible space. Note that since $\phi(\Omega) \subset \langle \phi(\Omega) \rangle$ then $\langle \phi(\Omega) \rangle$ is the smallest subspace containing all admissible vectors. Generally speaking, $|\langle \phi(\Omega) \rangle| >> |\phi(\Omega)|$.

Note that the regular representation (see Subsection 1.1) can be considered as a space embedding $\phi : \Omega \to W$ with respect to the group $G = \mathrm{Sym}(\Omega)$, where $\dim(W) = |\Omega| = n$ and $\phi : \omega \mapsto b_\omega$ with $\{b_\omega\}_{\omega \in \Omega}$ a basis of W. Also, $W = \langle \phi(\Omega) \rangle$.

A space embedding permits to construct a faithful representation of G, as explained in the next proposition.

Proposition 3.2. *Let $\alpha : \Omega \to W$ be a space embedding with respect to G. Suppose that $\forall \sigma \in G \quad \exists! A_\sigma \in \mathrm{GL}(W) \quad s.t. \quad \phi \circ \sigma = A_\sigma \circ \phi$. Then*

(1) we can define a map $\tilde{\phi} : G \to \mathrm{GL}(W)$, where $\tilde{\phi}(\sigma) = A_\sigma$, for any $\sigma \in G$;

(2) $\tilde{\phi}$ is a group homomorphism.

Proof. 1. Obvious.
2. We have to prove that $\tilde{\phi}(\sigma\sigma') = \tilde{\phi}(\sigma)\tilde{\phi}(\sigma')$ for all $\sigma, \sigma' \in G$, i.e. $A_{\sigma\sigma'} = A_\sigma A_{\sigma'}$. Using Definition 3.1, the following equality holds

$$A_{\sigma\sigma'}(\phi(\omega)) = \phi((\sigma\sigma')(\omega)) = \phi(\sigma(\sigma'(\omega))).$$

Since

$$A_\sigma A_{\sigma'}(\phi(\omega)) = A_\sigma(\phi(\sigma'(\omega))) = \phi(\sigma(\sigma'(\omega))),$$

we conclude that $A_{\sigma\sigma'} = A_\sigma A_{\sigma'}$, for all $\omega \in \Omega$. □

Remark 3.3. In Definition 3.1 we require only that A_σ exists, however in Theorem 3.2 we see that it is also unique.

For example, for the regular representation any permutation $\sigma \in \mathrm{Sym}(\Omega)$ defines a permutation $\sigma \in \mathrm{Sym}(\{b_\omega\}_{\omega\in\Omega})$ and so it defines a unique $A_\sigma \in \mathrm{GL}(W)$, which can be represented as a permutation matrix.

Now, we are interested in a special case of space embedding where the set Ω is a vector space $V = (\mathbb{F}_2)^r$ and W is the vector space $(\mathbb{F}_2)^s$, with $s > r$. For any $1 \le i \le s$, let $\mathbf{e}_i \in W$:

$$\mathbf{e}_i = (0, \ldots, 0, \underset{\substack{\uparrow \\ i}}{1}, 0, \ldots, 0)\,.$$

Let $\sigma \in \mathrm{Sym}(V)$ be any permutation over $(\mathbb{F}_2)^r$. We want to embed V into W by an injective map α and to extend σ to a permutation $\sigma' \in \mathrm{Sym}(W)$ as shown in the following commutative diagram:

$$\begin{array}{ccc} V & \xrightarrow{\alpha} & W \\ \downarrow{\scriptstyle\sigma} & \circlearrowleft & \downarrow{\scriptstyle\sigma'} \\ V & \xrightarrow{\alpha} & W \end{array}$$

In order to do this, we have to define the permutation $\sigma' \in \mathrm{Sym}(W)$. We say that σ' is an extension of σ. We seek a σ' that is linear on W. The following definition will be useful:

Definition 3.4. *Let $\sigma \in \mathrm{Sym}(V)$ and α be an injective map $\alpha : V \to W$. We say that σ is* **linearly extendible** *(via α) if $\forall\{v^i\}_{i\in I} \subset V$ we have*

$$\sum_{i\in I} \alpha(v^i) = 0 \iff \sum_{i\in I} \alpha(\sigma(v^i)) = 0.$$

Remark 3.5. Since we are considering the finite field $\mathbb{F}_2$, we note that σ is linearly extendible (via α) if $\forall\{v^i\}_{i\in I} \subset V$ such that $\sum_{i\in I} \alpha(v^i) = 0$ we have $\sum_{i\in I} \alpha(\sigma(v^i)) = 0$. In fact, an injective map defined on the set

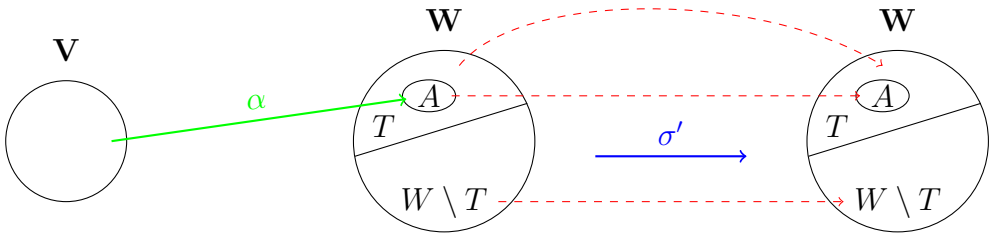
$$\{\{v^i\}_I \subset V \mid \sum_{i\in I} \alpha(v^i) = 0\}\}$$

into the set

$$\{\{\sigma(v^i)\}_I \subset V \mid \sum_{i\in I} \alpha(\sigma(v^i)) = 0\}\}$$

is a bijective map, since the cardinality of the two finite sets is the same.

Let $\alpha : V \to W$ be a space embedding. Let $A = \mathrm{Im}(\alpha) = \alpha(V)$ and let $T = \langle A\rangle$ be the subspace (the admissible space) of W linearly generated by A. Since $\sigma'(\alpha(v)) = \alpha(\sigma(v))$, $\forall v \in V$, we require that $\sigma'(A) = A$.



In order to specify the behavior of σ' on $(T \setminus A)$, which is the space of non-admissible vectors in the admissible space, we have to consider two different cases:

(a) suppose that σ is linearly extendible. Let $t \in T$, we must have $t = \sum_{1\le j\le \iota} a^j$, with $\iota \ge 1$, with $\{a^j\}_{1\le j\le i} \subset A$, $a^j = \alpha(v^j)$ (with $1 \le j \le \iota$ and $v^j \in V$). Then we define

$$\sigma'(t) = \sum_{1\le j\le \iota} \sigma'(a^j) = \sum_{1\le j\le \iota} \alpha(\sigma(v^j));$$

(b) in case σ is not linearly extendible, we define $\sigma'_{|_{T\setminus A}} = \mathrm{id}_{T\setminus A}$.

We now define σ' on $W \setminus T$ according to the two previous cases (i.e. depending on the behavior of σ on A).
In case (a), let τ be the dimension of the subspace T. We consider any subset B of $\{\mathbf{e}_1, \ldots, \mathbf{e}_s\}$ such that $|B| = s - \tau$ and W is the direct sum $W = T \oplus \langle B\rangle$. It is obvious that B exists. Let $w \in W$, then $w = w_T + w_B$ with $w_T \in T$ and $w_B \in \langle B\rangle$. Finally, we define

$$\sigma'(w) = \sigma'(w_T) + w_B.$$

In case (b) we define $\sigma'_{|_{W\setminus T}} = \mathrm{id}_{W\setminus T}$.

Lemma 3.6. *If σ is linearly extendible, then $\sigma' \in \mathrm{GL}(W)$.*

Proof. We first show that σ' is well-defined on T. Let $t = \sum_I a^i$ and $t' = \sum_J a^j$ and suppose that $t = t'$. Since σ is linearly extendible, we have

$$\begin{aligned} 0 = t + t' &= \sum_I a^i + \sum_J a^j = \sum_I \alpha(v^i) + \sum_J \alpha(v^j) = \sum_{I\cup J} \alpha(v^i) \\ \sigma'(t) + \sigma'(t') &= \sum_I \sigma'(a^i) + \sum_J \sigma'(a^j) = \sum_I \alpha(\sigma(v^i)) + \sum_J \alpha(\sigma(v^j)) \\ &= \sum_{I\cup J} \alpha(\sigma(v^i)) = 0. \end{aligned}$$

We now show that σ' is linear on T. Let $t_i = \sum_h a_h^{(i)}$. We have to show that $\sigma'(\sum_i t_i) = \sum_i \sigma'(t_i)$. Clearly,

$$\begin{aligned} \sigma'\big(\sum_i \sum_h a_h^{(i)}\big) = \sigma'\Big(\sum_{i,h} a_h^{(i)}\Big) &= \sum_i \sum_h \sigma'(a_h^{(i)}) = \sum_i \Big(\sum_h \sigma'(a_h^{(i)})\Big) \\ &= \sum_i \sigma'(t_i) \end{aligned}$$

and we have our thesis.
Since σ' is linear on T and T is a finite set, in order to prove that σ' is bijective on T it suffices to show that $\ker \sigma' = 0$. We have (by definition of linearly extendible)

$$0 = \sigma'(t) = \sum \alpha(\sigma(v^j)) \iff 0 = \sum \alpha(v^j) = t$$

Finally, we show the linearity on W. Let $\{w^i\}_{i\in I} \subset W$, we have to show the following equality

$$\sigma'\Big(\sum_{i\in I} w^i\Big) = \sum_{i\in I} \sigma'(w^i). \tag{1}$$

Since W is direct sum of T and $\langle B\rangle$, each element w^i in W can be considered as $w_T^i + w_B^i$ and so we can write the following

$$\begin{aligned} \sigma'\Big(\sum_{i\in I} w^i\Big) = \sigma'\Big(\sum_{i\in I} (w_T^i + w_B^i)\Big) &= \sigma'\Big(\sum_{i\in I} w_T^i\Big) + \sum_{i\in I} w_B^i \\ \sum_{i\in I} \sigma'(w^i) = \sum_{i\in I} \sigma'(w_T^i + w_B^i) &= \sum_{i\in I} \sigma'(w_T^i) + \sum_{i\in I} w_B^i. \end{aligned}$$

It easily follows that (1) holds if and only if

$$\sigma'\Big(\sum_{i\in I} w_T^i\Big) = \sum_{i\in I} \sigma'(w_T^i).$$

□

Remark 3.7. The construction of $\sigma' \in \mathrm{GL}(W)$ from σ linearly extendible (Definition 3.4) can be done similarly over any field.

We are now able to prove the main result of this subsection.

Theorem 3.8. *Let $W = (\mathbb{F}_2)^r$ and $G \leq \mathrm{Sym}(V)$. An injective map $\alpha : V \to W$ is a space embedding with respect to G if and only if, $\forall \sigma \in G$, σ is linearly extendible.*

Proof. Let α be a space embedding with respect to G. For any fixed $\sigma \in G$, there exists a map $A_\sigma \in \mathrm{GL}(W)$ such that $\alpha \circ \sigma = A_\sigma \circ \alpha$. Now, let $\{w^i\}_{i \in I}$ be a finite set such that $w^i = \alpha(v^i)$ (for any $i \in I$) and $\sum_{i \in I} w^i = 0$. Obviously we have

$$\sum_{i \in I} \alpha(\sigma(v^i)) = \sum_{i \in I} A_\sigma(\alpha(v^i)) = \sum_{i \in I} A_\sigma(w^i) = A_\sigma\Big(\sum_{i \in I} w^i\Big) = 0.$$

The converse immediately follows thanks to the previous lemma. □

Remark 3.9. For a fixed α and σ, the map σ' is unique and $\tilde{\alpha} : G \to \mathrm{GL}(W)$ is a representation of G, by Proposition 3.2.
Remark 3.10. In the following we use A_σ and σ' interchangeably.

3.2 A first embedding

We now apply the theory developed in the previous section to a specific space embedding[8] $\varepsilon : V \to W$.

Let us identify $(\mathbb{F}_2)^m$ with the field $\mathbb{F}_{2^m}$, via the quotient map $\mathbb{F}_{2^m} \leftrightarrow \mathbb{F}_2[x]/\langle \mathsf{p} \rangle$, where $\mathsf{p} \in \mathbb{F}_2[x]$ is any primitive polynomial such that $\deg(\mathsf{p}) = m$. We define a map $\varepsilon' : \mathbb{F}_{2^m} \to (\mathbb{F}_2)^{2^m}$ by means of a primitive element γ of $\mathbb{F}_{2^m}$ (which is a root of p). The map ε' is defined as

$$\varepsilon'(0) = (1, \underbrace{0, \ldots, 0}_{2^m - 1}) \qquad \varepsilon'(\gamma^i) = (0, \ldots, 0, \underset{\substack{\uparrow \\ i+1}}{1}, 0, \ldots, 0) \quad \forall 1 \leq i \leq 2^m - 1\,.$$

Note that $\varepsilon'(1) = \varepsilon'(\gamma^{2^m - 1}) = (\underbrace{0, \ldots, 0}_{2^m - 1}, 1)$.
Let b be a positive integer, let $r = mb$ and $s = 2^m b$. Let $V = (\mathbb{F}_2)^r$ and $W = (\mathbb{F}_2)^s$. We construct our injective map $\varepsilon : V \to W$ in the following way:

$$\varepsilon(v_1, \ldots, v_b) = (\varepsilon'(v_1), \ldots, \varepsilon'(v_b)) \tag{2}$$

for any $v_j \in (\mathbb{F}_2)^m$ ($1 \leq j \leq b$). Note that ε is a parallel[9] map.

For simplicity of notation, we set $e_1 = \varepsilon'(0) = (1, \underbrace{0, \ldots, 0}_{2^m - 1})$ and $e_{i+1} = \varepsilon'(\gamma^i)$, for any $1 \leq i \leq 2^m - 1$.

[8] which is called "α" in Subsection 3.1.
[9] see Subsection 1.4.

We note that

Lemma 3.11. *Suppose that* $\sum_{i\in I} e_i = e_h$. *Then* $h \in I$.

Proof. It follows from $\mathrm{w}(e_i) = 1$, for all $i \in I$. □

The following lemma is easily proved:

Lemma 3.12. *Let* I *be a finite index multiset such that* $\{v^i\}_I \subset V$. *For any* $1 \le h \le b$ *we have* $\sum_{i\in I} \varepsilon'(v_h^i) = 0$ *if and only if,* $\forall i \in I$, $|\{j \in I \mid v_h^j = v_h^i\}|$ *is even.*

Proof. Since ε' maps each element of $(\mathbb{F}_2)^m$ into the canonical basis of $(\mathbb{F}_2)^{2^m}$, each $\varepsilon'(v_h^i)$ is a vector such that $\mathrm{w}(\varepsilon'(v_h^i)) = 1$. Considering the following sum in $\mathbb{F}_2$, we have that $\sum_I \varepsilon'(v_h^i) = 0$ if and only if each component is made by an even number of 1, i.e. if and only if each element of the canonic basis that appears in our sum has an even weight. Since ε' is bijective, we have that $|\{j \in I \mid v_h^j = v_h^i\}|$ is even, $\forall i \in I$. □

Proposition 3.13. *Let* ε *as in (2). Then* $\dim_{\mathbb{F}_2}\big(\langle \mathrm{Im}(\varepsilon)\rangle\big) = 2^m b - (b-1)$.

Proof. We define the elements

$$z_{i,j} = (e_1, \ldots, e_1, \underset{\substack{\uparrow \\ i}}{e_j}, e_1, \ldots, e_1),$$

for $1 \le i \le b$ and $1 \le j \le 2^m$. Note that $z_{i,j} \neq z_{h,\ell}$ for $(i,j) \neq (h,\ell)$, except for $z_{11} = z_{21} = \ldots = z_{b1}$. We consider the set $\mathcal{B} = \{z_{1,1}\} \cup \{z_{i,j}\}_{j\ge 2,\ 1\le i\le b}$. For instance, when $m = 2$ and $b = 2$, we have

$$\mathcal{B} = \{(e_1,e_1), (e_1,e_2), (e_1,e_3), (e_1,e_4), (e_2,e_1), (e_3,e_1), (e_4,e_1)\}.$$

Clearly, the cardinality of the set $\mathcal{B}$ is given by

$$|\{z_{i,j}\}_{1\le b,\ 1\le j\le 2^m}| - |\{z_{i,1}\}_{i\ge 2}| = 2^m b - (b-1).$$

We claim that the set $\mathcal{B}$ is a basis for the subspace $\langle \mathrm{Im}(\varepsilon)\rangle$.
First, we prove that $\mathcal{B}$ is a linearly independent set. Suppose $z_{i,j} \in \mathcal{B}$ such that $(i,j) \neq (1,1)$. By definition of $\mathcal{B}$, the element $z_{i,j}$ is the unique element of $\mathcal{B}$ having a vector e_j in position i. Thus, $z_{i,j}$ cannot be the linear combination (i.e. a sum) of any other vectors of $\mathcal{B}$ (see Lemma 3.11). Now, we have to consider the element $z_{1,1}$. Let $z_{1,1} = \sum_{(i,j)\in J} z_{i,j}$. W.l.o.g., we can assume by Lemma 3.11 that there is $(\bar{i},\bar{j}) \in J$ such that $z_{\bar{i},\bar{j}} = (e_1, \ldots)$. Since $z_{\bar{i},\bar{j}} \neq z_{1,1}$ we can assume w.l.o.g. $z_{\bar{i},\bar{j}} = (e_1, e_{\bar{j}}, e_1, \ldots, e_1)$, i.e. $\bar{i} = 2$. There is no other $z_{i,j}$ having $e_{\bar{j}}$ in the second position. Therefore, the sum $z_{1,1}$ should contain a 1 in component $m + \bar{j}$, which is impossible.

Next, we prove that $\mathcal{B}$ generates $\langle \mathrm{Im}(\varepsilon)\rangle$. To do that, it suffices to prove that every element of $\mathrm{Im}(\varepsilon)$ belongs to the subspace generated by $\mathcal{B}$. If we

consider an element $w = (e_{j_1}, \ldots, e_{j_b}) \in \mathrm{Im}(\varepsilon)$, we have

$$w = \begin{cases} z_{1,j_1} + \cdots + z_{b,j_b} & \text{if } b \text{ is odd,} \\ z_{1,j_1} + \cdots + z_{b,j_b} + z_{1,1} & \text{if } b \text{ is even,} \end{cases}$$

since

$$\frac{\begin{matrix} (e_{j_1},\, e_1, \ldots,\, e_1) + \\ b-1 \left\{ \begin{matrix} (\,e_1, e_{j_2}, \ldots,\, e_1) + \\ \vdots \\ (\,e_1, \ldots,\, e_1\,, e_{j_b}) = \end{matrix} \right. \end{matrix} \quad b \text{ odd}}{(e_{j_1}, e_{j_2}, \ldots, e_{j_b})} \qquad \frac{\begin{matrix} (\,e_1,\, e_1, \ldots,\, e_1) + \\ (e_{j_1},\, e_1, \ldots,\, e_1) + \\ b-1 \left\{ \begin{matrix} (\,e_1, e_{j_2}, \ldots,\, e_1) + \\ \vdots \\ (\,e_1, \ldots,\, e_1\,, e_{j_b}) = \end{matrix} \right. \end{matrix} \quad b \text{ even}}{(e_{j_1}, e_{j_2}, \ldots, e_{j_b})}$$

□

Let $\mathcal{A}$ be a subset of the plaintext set $\mathcal{M}$ such that $|\mathcal{A}| = \dim_{\mathbb{F}_2}(\langle \mathrm{Im}(\varepsilon) \rangle) = 2^m b - (b-1)$. Let $a^i \in \mathcal{A}$, $1 \leq i \leq |\mathcal{A}|$. We construct the $(|\mathcal{A}| \times 2^m b)$-matrix $\mathbf{H}$ such that the i-th row is the image of the parallel map ε applied to the plaintext $a^i \in \mathcal{A}$, for $i \in \{1, \cdots, |\mathcal{A}|\}$:

$$\mathbf{H} = \begin{pmatrix} \varepsilon(a^1) \\ \varepsilon(a^2) \\ \vdots \\ \varepsilon(a^{|\mathcal{A}|}) \end{pmatrix} = \begin{pmatrix} \varepsilon'(a_1^1) & \varepsilon'(a_2^1) & \cdots & \varepsilon'(a_b^1) \\ \varepsilon'(a_1^2) & \varepsilon'(a_2^2) & \cdots & \varepsilon'(a_b^2) \\ \vdots & \vdots & \vdots & \vdots \\ \varepsilon'(a_1^{|\mathcal{A}|}) & \varepsilon'(a_2^{|\mathcal{A}|}) & \cdots & \varepsilon'(a_b^{|\mathcal{A}|}) \end{pmatrix}. \tag{3}$$

We would like to determine the expected rank for such a matrix. Generally speaking, for a random $(t \times n)$-matrix with entries in the finite field $\mathbb{F}_q$, we can use the following well known results:

Theorem 3.14 ([MMM04])**.** *Let $t, k, n \in \mathbb{N} \setminus \{0\}$, where $k \leq n$ and $k \leq t$.*

(1) The number of ordered k-tuples of linearly independent vectors in $(\mathbb{F}_q)^n$ is

$$(q^n - 1)(q^n - q)(q^n - q^2) \cdots (q^n - q^{k-1}).$$

(2) The number of k-dimensional subspaces of $(\mathbb{F}_q)^n$ is given by the q-binomial coefficient

$$\binom{n}{k}_q = \frac{\prod_{0 \leq i \leq k-1}(q^n - q^i)}{\prod_{0 \leq i \leq k-1}(q^k - q^i)}.$$

(3) The number of $(t \times n)$-matrices of rank k with entries in $\mathbb{F}_q$ is given by the following formula

$$d_{k,t} = \binom{n}{k}_q \prod_{0 \le i \le k-1} (q^t - q^i).$$

We note that

$$\frac{\binom{n}{k-1}_q}{\binom{n}{k}_q} = \frac{q^k - 1}{q^{n-k+1} - 1}. \tag{4}$$

By using the previous theorem, the relation in (4) and observing that

$$\frac{d_{t-2,t}}{d_{t,t}} = \frac{d_{t-2,t}}{d_{t-1,t}} \frac{d_{t-1,t}}{d_{t,t}}$$

we immediately get the following corollary:

Corollary 3.15. *Let $q = 2$ and suppose $t < n$. We have the following relations:*

$$\begin{aligned} d_{t,t} &= (2^n - 1)(2^n - 2) \cdots (2^n - 2^{t-1}); \\ \frac{d_{t-1,t}}{d_{t,t}} &= \frac{(2^t - 1)}{(2^n - 2^{t-1})} < \frac{1}{2^{n-t-1}} \le 1; \\ \frac{d_{t-2,t}}{d_{t,t}} &= \frac{(2^t - 1)(2^{t-1} - 1)}{3(2^n - 2^{t-2})(2^n - 2^{t-1})}. \end{aligned}$$

Corollary 3.16. *Let $q = 2$ and suppose $t = n$. We have the following relations:*

$$\begin{aligned} d_{n,n} &= (2^n - 1)(2^n - 2) \cdots (2^n - 2^{n-1}); \\ \frac{d_{n-1,n}}{d_{n,n}} &= \frac{2^n - 1}{2^{n-1}} \approx 2 > 1; \\ \frac{d_{n-2,n}}{d_{n,n}} &= \frac{(2^n - 1)(2^{n-1} - 1)}{9 \cdot 2^{2n-3}}. \end{aligned}$$

In other words, the probability that a $(t \times n)$ random matrix $(t < n)$ with entries in $\mathbb{F}_2$ has rank exactly t is significantly greater than the probability of having rank equal to $t - 1$ or $t - 2$ or less. Instead, the probability that a square $(n \times n)$ random matrix has rank $n - 1$ is the greatest.

Remark 3.17. In theory, the previous theorem cannot be applied to our case because our construction imposes specific constraints, for example on the row-weight. However, in practice our ratio $\frac{d_{t-1,t}}{d_{t,t}}$ approaches that of the Corollary 3.16 for $t = \dim_{\mathbb{F}_2}(\langle \mathrm{Im}(\varepsilon) \rangle)$.

So, in order to point out the distribution of the ranks of our matrices we provide a bound on the number of the full-rank matrices.

Lemma 3.18. *Let $c = 2^m$, let $n = cb$ $(n \geq k)$ and $z = \dim_{\mathbb{F}_2}\big(\langle \mathrm{Im}(\varepsilon)\rangle\big)$. The total number of admissible vectors in $\langle \mathrm{Im}(\varepsilon)\rangle$ is c^b. The average number $\xi(h)$ of admissible vectors in a subspace generated by h linearly independent admissible vectors is*

$$\xi(0) = 0, \quad \xi(1) = 1, \quad \xi(2) = 2,$$

$$\xi(h) = h + (2^h - h - 1)(\frac{c^b}{2^z}), \quad 3 \leq h \leq z - 1$$

Proof. An admissible vector can be any vector having weight 1 in any of the b components. There are c^b such vectors.

The whole space $\langle \mathrm{Im}(\varepsilon)\rangle$ contains 2^z vectors. The subspace B generated by h independent vectors $(V_1, \ldots, V_h)$ contains 2^h vectors. Of these, h are $(V_1, \ldots, V_h)$ themselves (admissible) and one is the zero vector (non-admissible).

So B contains $2^h - h - 1$ "other" vectors. To estimate how many of these are admissible, we simply multiply $2^h - h - 1$ by the ratio $\frac{\text{admissible vectors}}{\text{all vectors}} = \frac{c^b}{2^z}$. Therefore, our average contains $h + (2^h - h - 1)\frac{c^b}{2^z}$ admissible vectors □

Theorem 3.19. *Let $c = 2^m$, let $n = cb$ $(n \geq k \geq 1)$ and $z = \dim_{\mathbb{F}_2}\big(\langle \mathrm{Im}(\varepsilon)\rangle\big)$.*

(1) The number of $(k \times n)$-matrices having rank k can be estimated by the following formulas

$$\rho(k,k) = \prod_{1 \leq i \leq k} \big(c^b - \xi(i-1)\big)$$

i.e.

$$\rho(1,1) = c^b, \quad \rho(2,2) = c^b(c^b - 1), \quad \rho(3,3) = c^b(c^b - 1)(c^b - 2),$$

$$\rho(k,k) = c^b(c^b - 1)(c^b - 2) \prod_{4 \leq i \leq k} \left(c^b - (i-1) - (2^{i-1} - i)\frac{c^b}{2^z}\right), \quad k \geq 4$$

(2) The number of $(k \times n)$-matrices having rank $k-1$ can be estimated by the following recursive formula

$$\rho(2,1) = c^b, \quad \rho(3,2) = \rho(2,2)\xi(2) + \rho(2,1)(c^b - \xi(1)) = 3c^b(c^b - 1),$$

$$\rho(4,3) = \rho(3,3)\xi(3) + \rho(3,2)(c^b - \xi(2))$$

$$\rho(k,k-1) = \rho(k-1,k-1)\xi(k-1) + \rho(k-1,k-2)(2^z - 2^{k-2})\frac{c^b}{2^z}, \quad k \geq 5,$$

Proof. (1) In order for a $(k \times n)$-matrix to have rank k, the rows must be linearly independent. The first row can be any vector having weight 1 in any of the b component. There are c^b such vectors, so $\rho(1,1) = c^b$ (i.e. c^b is the total number of the admissible vectors). The second row must be independent of the first row. That means it cannot be equal

to the first row. There are $(c^b - 1)$ choices for the second row and thus $\rho(2,2) = c^b(c^b - 1)$. The third row cannot be equal to one of the previous rows. But also, in our representation, it is impossible that two admissible vectors add to another admissible vector. Then we have $(c^b - 2)$ choices for the third row, so $\rho(3,3) = c^b(c^b - 1)(c^b - 2)$.
On the other hand, if we add three or more admissible vectors we may get another admissible vector. As a consequence, if we are considering the i-th row, we must discard on average $\xi(i-1)$ vectors and so we can choose only among $c^b - \xi(i-1)$.

(2) The set of the $(k \times n)$ matrices having rank exactly $k-1$ is the disjoint union of two sets:

a) those having the first $k-1$ rows linearly independent (and so the k-th row dependent on the previous $k-1$ rows);
b) those having the first $k-1$ rows linearly dependent (and so these rows have rank $k-2$ and the k-th row is independent from them).

Therefore, the number of $(k \times n)$ matrices having rank exactly $k-1$ is obtained adding the following two values

a) the number of $(k-1) \times n$ matrices having rank $k-1$ multiplied by the number of all possible choices for the dependent row.

b) the number of $(k-1) \times n$ matrices having rank $k-2$ multiplied by the number of all possible choices for the independent row.

- The number of $(k-1) \times n$ matrices having rank $k-1$ is $\rho(k-1, k-1)$, for $k \geq 2$. In case $k = 2$, we have $\rho(1,1) = c^b$.
- The number of all possible choices for the dependent row is $\xi(k-1)$ for $k \geq 2$; if $k = 2$, the possible choice is exactly one, since the only second row we can choose is the first rows.
- The number of $(k-1) \times n$ matrices having rank $k-2$ is $\rho(k-1, k-2)$ and it makes sense for $k \geq 3$. When $k = 2$ we have to consider a matrix having exactly one row and with rank 0, so it is the *zero* row, but the *zero* row is not an admissible vector. In other words, when we have only two rows, the set in b) is empty. In case $k = 3$, we have $\rho(2,1) = c^b$, since the second row has to be equal to the first one.
- The number of all possible choices for the independent row is $(2^z - 2^{k-2})(\frac{c^b}{2^z})$ and it is true for $k \geq 5$. For $k = 3$, we must choose a third row different from the first two. The first two are equal and so we have $c^b - 1$ choices. For $k = 4$, we must choose a fourth row outside the space generated by the first three, but only two of the first three are distinct ans so we have $c^b - 2$ choices.

Putting altogether we obtain our formula.

□

3.3 Application to AES

Because of the AES structure, we assign the following values to the parameters we have previously introduced. Let $V = (\mathbb{F}_2)^r$ be our starting vector space with $r = 128$ and $W = (\mathbb{F}_2)^s$, $s > 128$. We need to establish s. We consider the quotient $\mathbb{F}_{256} \cong \mathbb{F}_2[x]/\langle \mathsf{m} \rangle$, where $\mathsf{m} = x^8+x^4+x^3+x+1 \in \mathbb{F}_2[x]$ is the AES-polynomial. So $m = 8$. According to the previous section, we consider $\varepsilon' : \mathbb{F}_{2^8} \to (\mathbb{F}_2)^{256}$ by means of a primitive element γ of $\mathbb{F}_{256}$, which is a root of the primitive polynomial[10] $\mathsf{n} = x^8 + x^4 + x^3 + x^2 + 1 \in \mathbb{F}_2[x]$, and we define our parallel map $\varepsilon : V \to W$, with $r = mb = 128$ and $s = 2^m b = 4096$, as

$$\varepsilon(v_1, \ldots, v_{16}) = (\varepsilon'(v_1), \ldots, \varepsilon'(v_{16})).$$

We have that $\dim_{\mathbb{F}_2} \big(\langle \mathrm{Im}(\varepsilon) \rangle\big) = 4081$, by Proposition 3.13.
A tipical round function of the AES cryptosystem consists of the composition of two parallel maps (AddRoundKey and SubBytes operations) and two non-parallel maps (ShiftRows and MixColumns operations). We view the SubBytes (and AddRoundKey) operation as a parallel map π

$$\begin{aligned} \pi : (\mathbb{F}_{2^8})^{16} &\to (\mathbb{F}_{2^8})^{16} \\ (y_1, \cdots, y_{16}) &\mapsto (\pi_1(y_1), \cdots, \pi_{16}(y_{16})) \end{aligned}$$

where $y_i \in \mathbb{F}_{2^8}$ and $\pi_i \in \mathrm{Sym}(\mathbb{F}_{256})$, for $1 \leq i \leq 16$. In the SubBytes case, each component π_i, where $1 \leq i \leq 16$, is composition of inversion operation and an affine map; in the AddRoundKey case, we have a sum with the round-key. By the Theorem 1.5 we recalled in the first section, we have that $\mathrm{Sym}(\mathbb{F}_{256}) = \langle ax+b, x^{254} \rangle$, where $a, b \in \mathbb{F}_{256}$. We note that a parallel map can be linearized using elementary results from Representation Theory.

Moreover, we claim that ShiftRows is linear over $(\mathbb{F}_2)^{4096}$ and that MixColumns is not linear over $(\mathbb{F}_2)^{4096}$, as follows.
First of all, we recall the map that describes the ShiftRows operation:

$$\begin{aligned} \mathsf{sh} : \quad (\mathbb{F}_{2^8})^{16} &\to (\mathbb{F}_{2^8})^{16} \\ (y_1, y_2, \cdots, y_{16}) &\mapsto (y_1, y_6, y_{11}, y_{16}, y_5, y_{10}, y_{15}, y_4, y_9, y_{14}, y_3, y_8, y_{13}, y_2, y_7, y_{12}). \end{aligned}$$

Denoting by $\mathbf{y} = (y_1, \cdots, y_{16})$, we note that

$$\varepsilon(\mathbf{y}) = (\varepsilon'(y_1), \varepsilon'(y_2), \varepsilon'(y_3), \varepsilon'(y_4), \varepsilon'(y_5), \cdots, \varepsilon'(y_{16}))$$

and

$$\varepsilon(\mathsf{sh}(\mathbf{y})) = (\varepsilon'(y_1), \varepsilon'(y_6), \varepsilon'(y_{11}), \varepsilon'(y_{16}), \varepsilon'(y_5), \cdots, \varepsilon'(y_{12})).$$

The map sh is linearly extendible because $\sum_{i \in I} \varepsilon(b^i) = 0$ clearly implies the following equality $\sum_{i \in I} \varepsilon(\mathsf{sh}(b^i)) = 0$.

[10] note that $\mathsf{n} \neq \mathsf{m}$; we could not use m because it is not primitive.

According to Lemma 3.6, it is possible to construct the linear map

$$A_{\text{sh}} : (\mathbb{F}_2)^{4096} \to (\mathbb{F}_2)^{4096}$$

and so the ShiftRows operation is linear over $(\mathbb{F}_2)^{4096}$.

Now, we show that the MixColumns operation is not linear over $(\mathbb{F}_2)^{4096}$ using the following counterexample.

Example 3.20. Let $w_1, w_2, w_3, w_4 \in W$ such that $w_1 + w_2 + w_3 = w_4$:

$$\begin{aligned}
w_1 &= (\varepsilon'(\gamma^1), \varepsilon'(\gamma^1), \varepsilon'(0), \varepsilon'(0), \varepsilon'(0), \cdots, \varepsilon'(0)) \\
w_2 &= (\varepsilon'(\gamma^1), \varepsilon'(0), \varepsilon'(\gamma^1), \varepsilon'(0), \varepsilon'(0), \cdots, \varepsilon'(0)) \\
w_3 &= (\varepsilon'(0), \varepsilon'(0), \varepsilon'(\gamma^1), \varepsilon'(0), \varepsilon'(0), \cdots, \varepsilon'(0)) \\
w_4 &= (\varepsilon'(0), \varepsilon'(\gamma^1), \varepsilon'(0), \varepsilon'(0), \varepsilon'(0), \cdots, \varepsilon'(0)).
\end{aligned}$$

Now, we apply the MixColumns operation MC to each vector w_1, w_2, w_3, w_4 obtaining the following

$$\begin{aligned}
\mathsf{MC}'(w_1) &= (\varepsilon'(\gamma^1), \varepsilon'(\gamma^3), \varepsilon'(0), \varepsilon'(\gamma^{51}), \varepsilon'(0), \cdots, \varepsilon'(0)) \\
\mathsf{MC}'(w_2) &= (\varepsilon'(\gamma^3), \varepsilon'(\gamma^{51}), \varepsilon'(\gamma^3), \varepsilon'(\gamma^{51}), \varepsilon'(0), \cdots, \varepsilon'(0)) \\
\mathsf{MC}'(w_3) &= (\varepsilon'(\gamma^1), \varepsilon'(\gamma^3), \varepsilon'(\gamma^{51}), \varepsilon'(\gamma^1), \varepsilon'(0), \cdots, \varepsilon'(0)) \\
\mathsf{MC}'(w_4) &= (\varepsilon'(\gamma^3), \varepsilon'(\gamma^{51}), \varepsilon'(\gamma^1), \varepsilon'(\gamma^1), \varepsilon'(0), \cdots, \varepsilon'(0))
\end{aligned}$$

where

$$\begin{array}{ccc} V & \xrightarrow{\varepsilon} & W \\ \downarrow{\scriptstyle \mathsf{MC}} & \circlearrowleft & \downarrow{\scriptstyle \mathsf{MC}'} \\ V & \xrightarrow{\varepsilon} & W \end{array} .$$

Then we have that $\mathsf{MC}'(w_1) + \mathsf{MC}'(w_2) + \mathsf{MC}'(w_3)$ is

$$(\varepsilon'(\gamma^3), \varepsilon'(\gamma^{51}), \varepsilon'(0) + \varepsilon'(\gamma^3) + \varepsilon'(\gamma^{51}), \varepsilon'(\gamma^1), \varepsilon'(0), \cdots, \varepsilon'(0)).$$

The third component of the previous vector is a sum in $(\mathbb{F}_2)^{256}$ and it has weight equal to 3. So, the vector $\mathsf{MC}'(w_1) + \mathsf{MC}'(w_2) + \mathsf{MC}'(w_3)$ is an element of the admissible space but it is a non-admissible vector.
Therefore, $\mathsf{MC}'(w_4) = \mathsf{MC}'(w_1 + w_2 + w_3) \neq \mathsf{MC}'(w_1) + \mathsf{MC}'(w_2) + \mathsf{MC}'(w_3)$ and so the MixColumns is not linear over W. It means that the extension of MC is not linearly extendible.

Remark 3.21. If all the AES operations were parallel maps, it would be possible to linearize the "full" cryptosystem because the set of the parallel maps is a group with respect to the composition operation.

3.4 Application to PRESENT

As for AES, we assign the right values to our parameters, according to PRESENT's structure. Let $V = (\mathbb{F}_2)^r$ be our starting vector space with $r = 64$, and $W = (\mathbb{F}_2)^s$ with $s > 64$. We consider $\varepsilon' : \mathbb{F}_{2^4} \to (\mathbb{F}_2)^{16}$ and we define our parallel map $\varepsilon : V \to W$, with $r = mb = 64$ and $s = 2^m b = 256$, as

$$\varepsilon(v_1, \ldots, v_{16}) = (\varepsilon'(v_1), \ldots, \varepsilon'(v_{16})).$$

We note that $\dim_{\mathbb{F}_2}\big(\langle \mathrm{Im}(\varepsilon)\rangle\big) = 241$ (see Proposition 3.13).
A typical round function of the PRESENT cryptosystem consists of the composition of two parallel maps (addRoundKey and sBoxLayer operations) and one non-parallel map (pLayer operation). The addRoundKey (and sBoxLayer) operation is a parallel maps π

$$\begin{aligned} \pi &: (\mathbb{F}_{2^4})^{16} \to (\mathbb{F}_{2^4})^{16} \\ (t_1, \cdots, t_{16}) &\mapsto (\pi_1(t_1), \cdots, \pi_{16}(t_{16})) \end{aligned}$$

where $\pi_i \in \mathrm{Sym}(\mathbb{F}_{16})$. In the sBoxLayer case, each component π_i $(1 \le i \le 16)$ is given by the table in Subsection 1.7; when π is the addRoundKey operation, we have only a bitwise sum with the round-key.
Moreover, it is easy to see that pLayer is not linear over $(\mathbb{F}_2)^{256}$.

Example 3.22. Let $w_1, w_2, w_3, w_4 \in W$ such that $w_1 + w_2 + w_3 = w_4$ and let $\zeta, \eta, \vartheta, \xi, \mu$ be distinct non-zero elements in $\mathbb{F}_{2^4}$. Suppose that

$$\begin{aligned} w_1 &= (\varepsilon'(\zeta), \varepsilon'(\zeta), \varepsilon'(0), \varepsilon'(0), \varepsilon'(0), \cdots, \varepsilon'(0)) \\ w_2 &= (\varepsilon'(\zeta), \varepsilon'(0), \varepsilon'(\zeta), \varepsilon'(0), \varepsilon'(0), \cdots, \varepsilon'(0)) \\ w_3 &= (\varepsilon'(0), \varepsilon'(0), \varepsilon'(\zeta), \varepsilon'(0), \varepsilon'(0), \cdots, \varepsilon'(0)) \\ w_4 &= (\varepsilon'(0), \varepsilon'(\zeta), \varepsilon'(0), \varepsilon'(0), \varepsilon'(0), \cdots, \varepsilon'(0)). \end{aligned}$$

Now, we apply the pLayer transformation pL to each vector w_1, w_2, w_3, w_4 obtaining the following

$$\begin{aligned} \mathsf{pL}'(w_1) &= (\varepsilon'(\eta), \varepsilon'(\mathbf{0})_3, \varepsilon'(\eta), \varepsilon'(\mathbf{0})_3, \varepsilon'(\eta), \varepsilon'(\mathbf{0})_3, \varepsilon'(\eta), \varepsilon'(\mathbf{0})_3) \\ \mathsf{pL}'(w_2) &= (\varepsilon'(\vartheta), \varepsilon'(\mathbf{0})_3, \varepsilon'(\vartheta), \varepsilon'(\mathbf{0})_3, \varepsilon'(\vartheta), \varepsilon'(\mathbf{0})_3, \varepsilon'(\vartheta), \varepsilon'(\mathbf{0})_3) \\ \mathsf{pL}'(w_3) &= (\varepsilon'(\xi), \varepsilon'(\mathbf{0})_3, \varepsilon'(\xi), \varepsilon'(\mathbf{0})_3, \varepsilon'(\xi), \varepsilon'(\mathbf{0})_3, \varepsilon'(\xi), \varepsilon'(\mathbf{0})_3) \\ \mathsf{pL}'(w_4) &= (\varepsilon'(\mu), \varepsilon'(\mathbf{0})_3, \varepsilon'(\mu), \varepsilon'(\mathbf{0})_3, \varepsilon'(\mu), \varepsilon'(\mathbf{0})_3, \varepsilon'(\mu), \varepsilon'(\mathbf{0})_3) \end{aligned}$$

where $\varepsilon'(\mathbf{0})_3$ means $(\varepsilon'(0), \varepsilon'(0), \varepsilon'(0))$. Then, we have that

$$\mathsf{pL}'(w_4) = \mathsf{pL}'(w_1+w_2+w_3) \neq \mathsf{pL}'(w_1)+\mathsf{pL}'(w_2)+\mathsf{pL}'(w_3) = (\varepsilon'(\eta)+\varepsilon'(\vartheta)+\varepsilon'(\xi), \ldots),$$

where the first component has weight 3, and so the pLayer is not a linear operation over W.

Remark 3.23. As in the AES case, if all the PRESENT's operations were parallel maps, it would be possible to linearize the "full" cryptosystem because the set of the parallel maps is a group with respect to the composition operation.

3.5 Application to SERPENT

Let $V = (\mathbb{F}_2)^r$ be our starting vector space with $r = 128$. In order to identify the value of $r \geq s$, where $W = (\mathbb{F}_2)^s$, we have to consider the map

$$\varepsilon' : (\mathbb{F}_{2^4}) \to (\mathbb{F}_2)^{2^4}.$$

We define our parallel map $\varepsilon : V \to W$ with $r = mb = 128$ and $s = 2^m b = 512$ as

$$\varepsilon(v_1, \ldots, v_{32}) = (\varepsilon'(v_1), \ldots, \varepsilon'(v_{32})).$$

Note that $\dim_{\mathbb{F}_2}(\langle \mathrm{Im}(\varepsilon) \rangle) = 2^m b - (b-1) = 481$.
The components of a typical round function are the parallel S-box, the affine transformation described in Subsection 1.6 and the translation with the round key. Obviously, key translation and S-box are parallel maps of type

$$\begin{aligned} \pi : (\mathbb{F}_{2^4})^{32} &\to (\mathbb{F}_{2^4})^{32} \\ (t_1, \ldots, t_{32}) &\mapsto (\pi_1(t_1), \ldots, \pi_{32}(t_{32})) \end{aligned}$$

where $\pi_i \in \mathrm{Sym}(\mathbb{F}_{2^4})$.

Similarly to what was done for AES and PRESENT, we could provide a counterexample to show that the linear transformation of SERPENT is not linear over $(\mathbb{F}_2)^{512}$.

4 Results on a larger embedding

In this section we provide another specific embedding that can be seen as an improvement of the former (2). Also the new embedding can be applied to AES, PRESENT and SERPENT. In Subsection 3.2 we considered $\Omega = V$ as a vector space and we found an embedding $V \hookrightarrow W$ such that the S-boxes and the key-additions become linear. However, in this way we lost the linearity of the Mixing Layer λ and so here we make a larger embedding where the linearity of λ is recovered, without losing the linearity of the key addition. We do lose the linearity of the S-boxes, but their non-linearity is probably kept low.

Starting from the setting we described in the previous section, we consider our parallel map $\varepsilon : (\mathbb{F}_{2^m})^b \to ((\mathbb{F}_2)^{2^m})^b$ defined as $\varepsilon(v_1, \ldots, v_b) = (\varepsilon'(v_1), \ldots, \varepsilon'(v_b))$.

Now, let $\mathbf{M}$ be a matrix in $\mathrm{GL}((\mathbb{F}_2)^{mb})$ and let t be its order, $\mathbf{M}^t = \mathrm{id}_V$. Let $V = (\mathbb{F}_2)^r$ be a vector space with dimension $r = mb$ and let $W = (\mathbb{F}_2)^s$ be the vector space with dimension $s = 2^m bt$. The space embedding $\alpha : V \to W$ we propose in this section is defined as follows

$$\alpha(v) = (\varepsilon(v), \varepsilon(\mathbf{M}v), \ldots, \varepsilon(\mathbf{M}^{t-1}v)). \tag{5}$$

From now on, α denotes the map in (5). Thanks to Proposition 3.13, we can easily prove the following proposition:

Proposition 4.1. *Let $V = (\mathbb{F}_2)^r$ be a vector space with dimension $r = mb$ and let $W = (\mathbb{F}_2)^s$ be the vector space with dimension $s = 2^m bt$. Let α be as in (5). Then we have*

$$2^m b - (b-1) \leq \dim_{\mathbb{F}_2} \big(\langle \mathrm{Im}(\alpha) \rangle\big) \leq (2^m b - (b-1))t$$

Proof. By Proposition 3.13, $\dim_{\mathbb{F}_2} \big(\langle \mathrm{Im}(\varepsilon) \rangle\big) = 2^m b - (b-1)$. Since

$$\{(\varepsilon(v), \varepsilon(\mathbf{M}v), \ldots, \varepsilon(\mathbf{M}^{t-1}v)) \mid v \in V\} \subset \{(\varepsilon(v_1), \ldots, \varepsilon(v_t)) \mid v_1, \ldots, v_t \in V\},$$

then

$$\dim_{\mathbb{F}_2} \big(\langle \mathrm{Im}(\alpha) \rangle\big) \leq (2^m b - (b-1))t.$$

On the other hand, considering the projection of $\{(\varepsilon(v), \varepsilon(\mathbf{M}v), \ldots, \varepsilon(\mathbf{M}^{t-1}v))\}$ on the first component (the first b bytes), the lower bound follows immediately, again considering Proposition 3.13. □

We can further improve Proposition 4.1 for byte-oriented Mixing Layer.

Proposition 4.2. *Let $V = (\mathbb{F}_2)^r$ be a vector space with dimension $r = mb$ and let $W = (\mathbb{F}_2)^s$ be the vector space with dimension $s = 2^m bt$. Let $\mathbf{M} \in \mathrm{GL}((\mathbb{F}_{2^m})^b)$. Let α be as in (5). Then we have*

$$\dim_{\mathbb{F}_2} \big(\langle \mathrm{Im}(\alpha) \big) \leq 2^m bt - (bt-1) - mb(t-1)$$

Proof. Let $T = \langle \mathrm{Im}(\alpha) \rangle$. For any $w_1, w_2 \in W$, let $w_1 \cdot w_2$ denote their scalar product. It is sufficient to show that there exist $(bt-1) + mb(t-1)$ elements in $T^\perp$ that are linearly independent, where $T^\perp = \{w \in W \mid w \cdot \mathbf{t} = 0, \forall \mathbf{t} \in T\}$ is the orthogonal space of T (or the "dual" of T, in coding theory notation). In fact, this means

$$\dim T^\perp \geq (bt-1) + mb(t-1)$$

and since $\dim T = \dim W - \dim T^\perp$, our result could follows.

Consider the following matrix product with $\mathbf{M} = (a_{i,j})$

$$\begin{pmatrix} a_{11} & a_{12} & \cdots & \cdots & a_{1b} \\ a_{21} & a_{22} & \cdots & \cdots & a_{2b} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ a_{b1} & a_{b2} & \cdots & \cdots & a_{bb} \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_b \end{pmatrix} = \begin{pmatrix} v'_1 \\ v'_2 \\ \vdots \\ v'_b \end{pmatrix}$$

Obviously, $v'_1 = \sum_{i=1}^{b} v_i a_{1i}$.
Let S' be a subspace of $(\mathbb{F}_2)^m$ such that $\dim(S') = m-1$. For any $1 \le i \le b$, let $S_i = \{\beta \in (\mathbb{F}_{2^m}) \mid \beta a_{1i} \in S'\}$. We note that S_i is a subspace and that

$$\Big\{ \sum_{i=1}^{b} v_i a_{1i} \mid v_i \in S_i,\, 1 \le i \le b \Big\} = S'$$

and that $|S_i| = |S'| = 2^{m-1}$. There exists a bijection via orthogonality between the sets $\mathcal{S} = \{S < (\mathbb{F}_2)^m | \dim(S) = m-1\}$ and $\{S^\perp < (\mathbb{F}_2)^m | \dim(S^\perp) = 1\}$; their cardinality is obviously $2^m - 1$. We can choose a linear basis for $\mathcal{S} \cup \{0\}$, i.e. $\mathcal{S} \cup \{0\} = \langle \mathbf{e}_1^\perp, \ldots, \mathbf{e}_m^\perp \rangle$. Therefore, each row of $\mathbf{M}$ generates m linearly independent elements of $T^\perp$.
Two relations coming from two different rows are independent, since the matrix $\mathbf{M}$ has full rank, for a total of mb relations.

Now, we construct the elements of the orthogonal space that correspond to the relations induced by the rows of $\mathbf{M}$. We are considering the case $(v, \mathbf{M}v)$ and we observe that

$$\sum_{i=1}^{b} v_i a_{1i} = v'_1 = (\mathbf{M}v)_1 \tag{6}$$

where $v_i \in S_i$. Since $\varepsilon'(S_i) \subset (\mathbb{F}_2)^{2^m}$, we consider $w_i = \sum_{\ell \in \varepsilon'(S_i)} \ell$ where $\mathrm{w}(w_i) = |\varepsilon'(S_i)| = 2^{m-1}$ and $w_i \in (\mathbb{F}_2)^{2^m}$. The element of $T^\perp$ coming from (6) and S is

$$(w_1, \ldots, w_b, w'_1, \ldots, 0, \ldots 0)$$

where $w'_1 = \sum_{\ell \in \varepsilon'(S')} \ell$. Clearly, $m-1$ similar elements come from (6) and $\mathcal{S}$. If we consider the relations given by the h-th row of $\mathbf{M}$, i.e. $\sum_{i=1}^{b} v_i a_{hi} = v'_h$, we obtain the following elements

$$(w_1, \ldots, w_b, 0, \ldots, w'_h, \ldots, 0, \ldots 0).$$

At this point, we have constructed the mb elements of the orthogonal space corresponding to the previous relations.
Instead of considering $(v, \mathbf{M}v)$, since clearly $\mathbf{M}(\mathbf{M}^i v) = \mathbf{M}^{i+1} v$, we can apply the previous construction to each pair $(\mathbf{M}^i v, \mathbf{M}^{i+1} v)$, for $1 \le i \le t-2$,

obtaining the corresponding elements

$$(\underbrace{0,\ldots,0}_{b(i-1)},\underbrace{w_1,\ldots,w_b}_{b},\underbrace{0,\ldots,w'_h,\ldots,0}_{b},\underbrace{0,\ldots,0}_{bt-(i+1)b}) \tag{7}$$

We have found exactly $mb(t-1)$ vectors in $T^\perp$. Since the pairs $(\mathbf{M}^i v, \mathbf{M}^{i+1}v)$ and $(\mathbf{M}^j v, \mathbf{M}^{j+1}v)$ with $i \neq j$ involve different bytes, the relations given by $(\mathbf{M}^i v, \mathbf{M}^{i+1}v)$ are independent from those given by $(\mathbf{M}^j v, \mathbf{M}^{j+1}v)$. Then we have $mb(t-1)$ independent relations (i.e. linearly independent elements of the orthogonal space).

Thanks to Proposition 3.13, we have exactly $(bt-1)$ further relations, corresponding to elements in $T^\perp$ of type

$$(\underbrace{0,\ldots,0}_{k-1},\underbrace{1,\ldots,1}_{b},\underbrace{1,\ldots,1}_{b},\underbrace{0,\ldots,0}_{bt-(k+1)}) \tag{8}$$

with $1 \leq k \leq bt$.

The vectors (7) and (8) form clearly a linearly independent set. □

As we have done in previous section, we can construct the following matrix. Let $\mathcal{D}$ be a subset of the plaintext set $\mathcal{M}$ such that $|\mathcal{D}| = \dim_{\mathbb{F}_2}(\langle \mathrm{Im}(\alpha)\rangle)$. Let $a_i \in \mathcal{D}$, $1 \leq i \leq |\mathcal{D}|$. We construct the $(|\mathcal{D}| \times 2^m bt)$-matrix $\mathbf{D}$ such that the i-th row is the image of the map α applied to the plaintext $a_i \in \mathcal{D}$, for $i \in \{1, \cdots, |\mathcal{D}|\}$:

$$\mathbf{D} = \begin{pmatrix} \alpha(a^1) \\ \alpha(a^2) \\ \vdots \\ \alpha(a^{|\mathcal{D}|}) \end{pmatrix} = \begin{pmatrix} \varepsilon(a^1) & \varepsilon(\mathbf{M}a^1) & \cdots & \varepsilon(\mathbf{M}^{t-1}a^1) \\ \varepsilon(a^2) & \varepsilon(\mathbf{M}a^2) & \cdots & \varepsilon(\mathbf{M}^{t-1}a^2) \\ \vdots & \vdots & \vdots & \vdots \\ \varepsilon(a^{|\mathcal{D}|}) & \varepsilon(\mathbf{M}a^{|\mathcal{D}|}) & \cdots & \varepsilon(M^{t-1}a^{|\mathcal{D}|}) \end{pmatrix}.$$

Remark 4.3. We expect the rank of this matrix to have a behavior similar to that of matrix $\mathbf{H}$ (3), see Remark 3.17. Our experiments confirm this.

Let $\tilde{\mathcal{G}}$ be the set of parallel maps $\tilde{\pi} : (\mathbb{F}_{2^m})^b \to (\mathbb{F}_{2^m})^b$, such that, for any $1 \leq j \leq b$, $\tilde{\pi}_j(x) = ax + c$, with $a \neq 0, c \in \mathbb{F}_{2^m}$ (a and c do not depend on j). Let $\bar{\mathcal{G}}$ be the set of parallel maps $\bar{\pi} : (\mathbb{F}_{2^m})^b \to (\mathbb{F}_{2^m})^b$, such that, for any $1 \leq j \leq b$, $\bar{\pi}_j(x) = x + d_j$, with $d_j \in \mathbb{F}_{2^m}$.
Note that both $\tilde{\mathcal{G}}$ and $\bar{\mathcal{G}}$ are subgroups of $\mathrm{Sym}((\mathbb{F}_{2^m})^b)$ and we define $\mathcal{G}$ as

$$\mathcal{G} = \left\langle \tilde{\mathcal{G}}, \bar{\mathcal{G}}, \mathbf{M} \right\rangle < \mathrm{Sym}((\mathbb{F}_{2^m})^b).$$

The following result holds:

Proposition 4.4. *Let σ be either an element of $\tilde{\mathcal{G}}$ or an element of $\bar{\mathcal{G}}$, then there exists $A_\sigma : W \to W$ which is linear.*

Proof. We want to apply Lemma 3.6 and so we must only show that σ is linearly extendible. Let $\{v^i\}_{i\in I} \subset V$ such that $\sum_{i\in I}\alpha(v^i)=0$, we have to prove that $\sum_{i\in I}\alpha(\sigma(v^i))=0$. Note that $\sum_I \alpha(v^i)=0$ is equivalent to

$$\sum_I \left(\varepsilon'(v^i)_1,\dots,\varepsilon'(v^i)_b,\varepsilon'(\mathbf{M}v^i)_1,\dots,\varepsilon'(\mathbf{M}v^i)_b,\dots,\varepsilon'(\mathbf{M}^{t-1}v^i)_1,\dots,\varepsilon'(\mathbf{M}^{t-1}v^i)_b\right)=0.$$

Then we have the following system S_j for any $1\le j\le b$

$$S_j=\begin{cases}\sum_I \varepsilon'(v^i_j)=0\\ \sum_I \varepsilon'((\mathbf{M}v^i)_j)=0\\ \vdots\\ \sum_I \varepsilon'((\mathbf{M}^{t-1}v^i)_j)=0.\end{cases}$$

Using Lemma 3.12, we have that S_j is equivalent to S'_j

$$S'_j=\begin{cases}|\{\ell \mid v^\ell_j=v^i_j\}| \text{ is even } \forall i\in I\\ |\{\ell \mid (\mathbf{M}v^\ell)_j=(\mathbf{M}v^i)_j\}| \text{ is even } \forall i\in I\\ \vdots\\ |\{\ell \mid (\mathbf{M}^{t-1}v^\ell)_j=(\mathbf{M}^{t-1}v^i)_j\}| \text{ is even } \forall i\in I.\end{cases}$$

Suppose $\sigma\in\tilde{\mathcal{G}}$ which means that $\sigma(v)=\sigma(v_1,\cdots,v_b)=(\sigma_1(v_1),\cdots,\sigma_b(v_b))$ where $\sigma_i(v_i)=av_i+c$ for any $1\le i\le b$ and $a\neq 0, c\in\mathbb{F}_{2^m}$.
Since $\mathbf{M}$ is linear, we have

$$\begin{aligned}(\mathbf{M}^h\sigma(v^\ell))_j&=(\mathbf{M}^h(av^\ell_1+c,\cdots,av^\ell_b+c))_j\\ &=(a\mathbf{M}^hv^\ell+\mathbf{M}^h(c,\cdots,c))_j\\ &=(a\mathbf{M}^hv^\ell)_j+(\mathbf{M}^h(c,\cdots,c))_j\\ &=a(\mathbf{M}^hv^\ell)_j+\bar{c},\end{aligned}$$

where $\bar{c}$ is a constant *independent* of ℓ.
We have that, $\forall i\in I$ and for any $1\le h\le t-1$, $|\{\ell \mid (\mathbf{M}^hv^\ell)_j=(\mathbf{M}^hv^i)_j\}|$ is even and so that $|\{\ell \mid a(\mathbf{M}^hv^\ell)_j+\bar{c}=a(\mathbf{M}^hv^i)_j+\bar{c}\}|$ is even. Thanks to Lemma 3.12, our thesis follows.

Suppose now that $\sigma\in\bar{\mathcal{G}}$, i.e. $\sigma(v)=v+d$ for some $d\in V$. Since

$$\begin{aligned}(\mathbf{M}^h\sigma(v^\ell))_j&=(\mathbf{M}^h(av^\ell+d))_j\\ &=(\mathbf{M}^hv^\ell)_j+(\mathbf{M}^h(d))_j\\ &=(\mathbf{M}^hv^\ell)_j+\bar{d},\end{aligned}$$

where $\bar{d}$ is a constant *independent* of ℓ and $|\{\ell \mid (\mathbf{M}^hv^\ell)_j=(\mathbf{M}^hv^i)_j\}|$ is even, we have that

$$|\{\ell \mid (\mathbf{M}^hv^\ell)_j+\bar{d}=(\mathbf{M}^hv^i)_j+\bar{d}\}|$$

is even. By Lemma 3.12, our thesis follows. □

4.1 Application to AES

Let $V = (\mathbb{F}_2)^r$ be a vector space with dimension $r = 128$ and let $\mathsf{M} : V \to V$ be the MixingLayer of AES, that is, the composition of ShiftRows and MixColumns. Since M has order equal to 8 (i.e. $\mathsf{M}^8 = \mathrm{id}_V$), the map $\alpha : V \to W$ we propose is defined as follows

$$\alpha(v) = (\varepsilon(v), \varepsilon(\mathsf{M}v), \ldots, \varepsilon(\mathsf{M}^7 v)), \tag{9}$$

where $W = (\mathbb{F}_2)^s$ is the vector space with dimension $s = 2^m bt = 2^{15}$ and ε is the map defined in Subsection 3.3: $\varepsilon : (\mathbb{F}_2)^{128} \to (\mathbb{F}_2)^{4096}$.
Let $T = \langle \mathrm{Im}(\alpha) \rangle$ with α in (9). We can easily determine $\dim(T)$.

Fact 4.5. *In the AES case we have*

$$\dim_{\mathbb{F}_2}(T) = 2^m bt - (bt - 1) - mb(t - 1) = 31745.$$

Proof. Let $\lambda = 2^m bt - (bt - 1) - mb(t - 1)$. By computational experiments, we have found a $(\lambda \times 2^m bt)$ full rank matrix for the α representation in the AES case. Which means $\dim_{\mathbb{F}_2} T \geq \lambda$. Thanks to Proposition 4.2 we conclude that $\dim_{\mathbb{F}_2} T = \lambda$. □

We note that the group

$$\mathcal{G} = \left\langle \tilde{\mathcal{G}}, \bar{\mathcal{G}}, \mathsf{M} \right\rangle < \mathrm{Sym}((\mathbb{F}_{2^8})^{16}).$$

contains all the permutations of the AES-round function, except notably for the S-box operation.

Proposition 4.6. *Let* M *be the* MixingLayer. *Then* α *is a space embedding with respect to* $\mathcal{G} = \left\langle \tilde{\mathcal{G}}, \bar{\mathcal{G}}, \mathsf{M} \right\rangle$.

Proof. According to Proposition 4.4, there exists a linear map $A_\sigma : W \to W$ in case σ is $\tilde{\mathcal{G}}$ or $\bar{\mathcal{G}}$. We note that the previous result is independent from $\mathbf{M}$. Let $\mathbf{M}$ be the MixingLayer M. Since $\alpha(v^i) = (\varepsilon(v^i), \varepsilon(\mathsf{M}v^i), \ldots, \varepsilon(\mathsf{M}^7 v^i))$ and

$$\begin{aligned} \alpha(\mathsf{M}v^i) &= (\varepsilon(\mathsf{M}v^i), \varepsilon(\mathsf{M}^2 v^i), \ldots, \varepsilon(\mathsf{M}^8 v^i)) \\ &= (\varepsilon(\mathsf{M}v^i), \varepsilon(\mathsf{M}^2 v^i), \ldots, \varepsilon(v^i)) \end{aligned}$$

$\alpha(\mathsf{M}v^i)$ is a permutation of $\alpha(v^i)$. Obviously, we have that $\sum_{i \in I} \alpha(v^i) = 0$ implies $\sum_{i \in I} \alpha(\mathsf{M}v^i) = 0$. □

With a fixed K, the encryption ϕ_K is the composition of AddRoundKey, Subbytes and MixingLayer. So the only part of ϕ_K which is not linear (with our map α) is the SubBytes operation.

4.2 Application to PRESENT

Let $V = (\mathbb{F}_2)^r$ be a vector space with dimension $r = 64$ and let $\mathsf{M} : V \to V$ be the pLayer of PRESENT. Since $\mathsf{M}^3 = \mathrm{id}_V$, the map $\alpha : V \to W$ we propose is defined as follows

$$\alpha(v) = (\varepsilon(v), \varepsilon(\mathsf{M}v), \varepsilon(\mathsf{M}^2 v)), \tag{10}$$

where $W = (\mathbb{F}_2)^s$ is the vector space with dimension $s = 2^m bt = 768$. Let α be as in (10) and $T = \langle \mathrm{Im}(\alpha) \rangle$. Also in this case it is possible to prove (with a computation) that $\dim_{\mathbb{F}_2}(T) = 2^m bt - (bt - 1) - mb(t - 1) = 593$

With a fixed K, the encryption ϕ_K is the composition of addRoundKey, sBoxLayer and pLayer. So the only part of ϕ_K which is not linear (with our map α) is the sBoxlayer operation.

4.3 Application to SERPENT

Let $V = (\mathbb{F}_2)^r$ be a vector space with dimension $r = 128$ and let $\mathsf{M} : V \to V$ be the affine transformation of SERPENT. Since the order of M is greater [11] than 2^{116}, it is huge and impractical to consider the map $\alpha : V \to W$

$$\alpha(v) = (\varepsilon(v), \varepsilon(\mathsf{M}v), \ldots, \varepsilon(\mathsf{M}^{2^{80}} v), \ldots). \tag{11}$$

since $W = (\mathbb{F}_2)^s$ would have $s = 2^m bt > 2^4 \cdot 32 \cdot 2^{116} = 2^{125}$, making the rank computation impossible with nowadays technology.

5 Attack strategies

In this paper we do not report on successful attacks on (full versions of) the AES or other well-known ciphers. It is true that we have implemented several attacks aiming at distinguishing AES from random permutations, presented in some talks, and that we have collected some data indicating that our approach is likely to succeed. Yet, our data do not provide an overwhelming statistical evidence for the full cipher versions. Therefore, in this section we sketch some attack strategies that we have followed, without giving full details.

The most difficult task in assessing the success of one of our embeddings is, by far, to estimate the non-linearity decrease of the cryptosystem. For example, a rigorous determination of the s-extendibility (Subsection 6.1) appears completely out of reach. The only methods we can use to estimate the non-linearity fall are "a posteriori" checks on linear dependences.

We have implemented only chosen-plaintex attacks, either with single-key or with related keys. In the single-key scenario, we proceed in three steps:

[11] to be precise it is 110329570561973845861261474090270635, as computed directly with MAGMA.

(1) we choose a set S of N (31745×2^{15})-matrices, with rows taken from T (Fact 4.5);

(2) we encrypt all matrices in S (row by row) with a given key and compute their ranks;

(3) we compare their rank distribution with the expected rank distribution for a set of N **random** (31745×2^{15})-matrices, with rows taken from T, aiming at distinguishing the two distributions;

(4) to validate the distinguishing statistical test, we also create sets of N **random** (31745×2^{15})-matrices (in T) and we compare them with the expected distribution, aiming at *not* distinguishing them.

In the related-key scenario we proceed similarly. Let n_k be the number of related keys:

(1) we choose a set S of N (31745×2^{15})-matrices, with rows taken from T;

(2) we encrypt all matrices in S (row by row) with all keys and compute their ranks;

(3) we compare their rank distribution with the expected rank distribution for a set of NN_k **random** (31745×2^{15})-matrices, with rows taken from T, aiming at distinguishing the two distributions;

(4) to validate the distinguishing statistical test, we also create sets of Nn_k **random** (31745×2^{15})-matrices (in T) and we compare them with the expected distribution, aiming at *not* distinguishing them.

Apart from the obvious difference in the dealing of the single-key/related-key mechanism, the two scenarios are very similar, since in both we hope to spot a significant deviation by looking at ranks. Matrix ranks do depend on the linear dependences of the rows and are much easier to compute and compare, so they are cheap indicators for the non-linearity behavior (see Marsaglia's test, e.g. [Sot98],[NIS00]).
On the other hands, since a great deal of row dependences influence the rank, as indicators they are noisy and force us to collect a huge number of samples. To maximize the effect on the rank of our embeddings, we need to choose S with a very specific rank distribution, e.g. with matrices of extremely low rank (while keeping all rows distinct).
A report on some experimental results can be found in [RSB10].

6 Further remarks and other results

The first subsection contains some results on how our representation could achieve a weaker notion of linearity.

In Subsection 6.2 we report other thinkable representations, that unfortu-

nately are impractical. The main objective in these constructions is to identify the right compromise between computational feasibility and quantity of information that can be obtained.

Then, in Subsection 6.3 we prove the fact, using classical and easy arguments, that it is unlikely to embed the AES cipher into a linear cipher, unless one uses a huge-dimensional vector space (and so this embedding is useless in practice).

6.1 On a weaker notion of linearity

The results in this section are jointly with L. Maines and the proofs are contained in her Master's thesis [Mai09] (see also [MRS10]), supervised by the second author.
The main goal sought in Section 3.1, Section 3.2, Section 4, and Section 6.2 is to find practical embedding of $(\mathbb{F}_2)^{128}$ into a larger space where all components of the round function become linear. This is impossible, as shown in Section 6.3, but what we achieve in Section 4 is an embedding where the non-linear maps are "not so far" from linear maps. There are many notions of "non-linearity", but none of them can be easily computed in our setting. When we say "not so far from linear", we mean that these functions behave with matrix ranks in a way similar to that of linear maps, as discussed in Section 5.

However, we have been able to introduce a new non -linearity notion, that we call *s-extendibility* (Definition 6.1). We are not able to apply it in the embedding

$$\alpha : v \to (\varepsilon(v), \varepsilon(\mathbf{M}v), \cdots, \varepsilon(\mathbf{M}^7 v)). \tag{12}$$

but we can apply it [12] to

$$\alpha : v \to (\varepsilon(v), \varepsilon(\mathbf{M}v)).$$

and so our definition and our results (the main results of this section is Theorem 6.6) should be seen as a step forward the complete understanding of the surviving non-linearity in (12).

Definition 6.1. *Let $V = (\mathbb{F}_2)^r$ and $W = (\mathbb{F}_2)^s$, with $s > r$. Let $\sigma \in \mathrm{Sym}(V)$ and α be an injective map $\alpha : V \to W$. We say that σ is* **s-extendible** *(via α) if $\forall \{v^h\}_{1 \le h \le s} \subset V$ we have*

$$\sum_{h=1}^{s} \alpha(v^h) = 0 \iff \sum_{h=1}^{s} \alpha(\sigma(v^h)) = 0.$$

Remark 6.2. If $v^1 = v^2$ and $v^3 = v^4$, then $\forall \alpha$ and $\forall \sigma$ we have

$$\alpha(v^i) + \alpha(v^2) + \alpha(v^3) + \alpha(v^4) = 0$$

[12] under specific conditions on M

and

$$\alpha(\sigma(v^i)) + \alpha(\sigma(v^2)) + \alpha(\sigma(v^3)) + \alpha(\sigma(v^4)) = 0.$$

So if we test the 4-extendibility of σ only on these sets of vectors, we will find that any σ is 4-extendible. We call these vectors "coupled vectors".

We note that if σ is s-extendible $\forall s \in \mathbf{N}$, then σ is linearly extendible, according to Definition 3.4. Moreover, any linear map is s-extendible for all s. A random map is a 2-extendible but (with high probability) it is not s-extendible for any $s \geq 4$. Therefore, any 4-extendible map can be considered closer to a linear map. We would like to have results on our embedding concerning the s-extendibility of maps. A first result in this direction is obtained using the space embedding

$$\alpha(v) = (\varepsilon(v), \varepsilon(\mathbf{M}v)), \tag{13}$$

where $\mathbf{M}$ is a $(n \times n)$-matrix with entries in $\mathbb{F}_{2^m}$, as we are going to explain.

Definition 6.3. *Let $i, j, x, y, \alpha, \beta, \cdots \in \mathbb{F}_{2^m}$ and $\mathbf{M}$ an $(n \times n)$-matrix with entries in $\mathbb{F}_{2^m}$*

$$\mathbf{M} = \begin{pmatrix} m_{11} & m_{12} & \dots & m_{1n} \\ m_{21} & m_{22} & \dots & \dots \\ \vdots & \vdots & \ddots & \ddots \\ m_{n1} & \dots & \dots & m_{nn} \end{pmatrix}.$$

The vectors $w_1, w_2, w_3, w_4 \in (\mathbb{F}_{2^m})^{2n}$ are 4-related vectors *if they can be permuted in order to have the following form:*

	1	2	$\dots$	$n+1$	$\dots$	$2n$
1 w_1,	$(i,$	$x,$	$\dots$	$m_{11}i + m_{12}x + \dots,$	$\dots$	$m_{n1}i + m_{n2}x + \dots)$
2 w_2,	$(i,$	$y,$	$\dots$	$m_{11}i + m_{12}y + \dots,$	$\dots$	$m_{n1}i + m_{n2}y + \dots)$
3 w_3,	$(j,$	$x,$	$\dots$	$m_{11}j + m_{12}x + \dots,$	$\dots$	$m_{n1}j + m_{n2}x + \dots)$
4 w_4,	$(j,$	$y,$	$\dots$	$m_{11}j + m_{12}y + \dots,$	$\dots$	$m_{n1}j + m_{n2}y + \dots)$

Four related vectors $w_1, \dots, w_4$ are admissible vectors $\alpha(v_1) = (\varepsilon(v_1), \varepsilon(\mathbf{M}v_1))$, $\alpha(v_2) = (\varepsilon(v_2), \varepsilon(\mathbf{M}v_2))$, $\alpha(v_3) = (\varepsilon(v_3), \varepsilon(\mathbf{M}v_3))$, $\alpha(v_4) = (\varepsilon(v_4), \varepsilon(\mathbf{M}v_4))$ such that

$$\varepsilon(v_1) + \varepsilon(v_2) + \varepsilon(v_3) + \varepsilon(v_4) = 0\,,$$

but we do not know the sum $\varepsilon(\mathbf{M}v_1) + \varepsilon(\mathbf{M}v_2) + \varepsilon(\mathbf{M}v_3) + \varepsilon(\mathbf{M}v_4)$.

Let σ be a parallel maps over $(\mathbb{F}_{2^m})^{2n}$. The image of 4-related vectors via σ can be seen as

	1	2	...	$n+1$	...	$2n$
1	$w_1,$	$(\sigma(i),$	$\sigma(x),$ ...	$m_{11}\sigma(i)+m_{12}\sigma(x)+\ldots,$	...	$m_{n1}\sigma(i)+m_{n2}\sigma(x)+\ldots)$
2	$w_2^*,$	$(\sigma(i),$	$\sigma(y),$ ...	$m_{11}\sigma(i)+m_{12}\sigma(y)+\ldots,$	...	$m_{n1}\sigma(i)+m_{n2}\sigma(y)+\ldots)$
3	$w_3^*,$	$(\sigma(j),$	$\sigma(x),$ ...	$m_{11}\sigma(j)+m_{12}\sigma(x)+\ldots,$	...	$m_{n1}\sigma(j)+m_{n2}\sigma(x)+\ldots)$
4	$w_4^*,$	$(\sigma(j),$	$\sigma(y),$ ...	$m_{11}\sigma(j)+m_{12}\sigma(y)+\ldots,$	...	$m_{n1}\sigma(j)+m_{n2}\sigma(y)+\ldots)$

Definition 6.4. *4-related vectors $w_1,\ldots,w_4$ are* totally *4-related if*

$$w_1+w_2+w_3+w_4=0.$$

Definition 6.5. *Given $(x,y,z,a,b,c)\in\mathbb{N}^6$ and an $(n\times n)$-matrix $\mathbf{M}$, we say that (x,y,z,a,b,c)* fits $\mathbf{M}$ *if the following sums of elements of* $\det(\mathbf{M})$ *are non-zero:*

- *the sums having a number of elements equal to*

$$\sum_{i=0}^{x}\binom{n-c}{i}\binom{n-b}{x-i}\binom{b-i}{y}x!y! \quad \text{when } z=0, x\neq 0, y\neq 0$$

$$\sum_{i=0}^{x}\binom{n-b}{i}\binom{n-c}{x-i}\binom{c-i}{z}x!z! \quad \text{when } y=0, x\neq 0, z\neq 0$$

$$\sum_{i=0}^{y}\binom{n-a}{i}\binom{n-c}{x-i}\binom{c-i}{z}y!z! \quad \text{when } x=0, y\neq 0, z\neq 0$$

- *the sums having a number of elements equal to*

$$\sum_{i=0}^{x}\sum_{j=0}^{y}\binom{n-c}{i}\binom{n-b}{x-i}\binom{n-a}{j}\binom{(n-c)-i}{y-j}\binom{c-(x-i)-j}{z}x!y!z!$$

when $x\neq 0, y\neq 0, z\neq 0$.

The main result of this section is the next theorem that gives sufficient conditions on $\mathbf{M}$ in order to make all $\sigma: V\to V$ into 4-exendible maps.

Theorem 6.6. *Let $\mathbf{M}$ be an $(n\times n)$-matrix, with entries in $\mathbb{F}_{2^m}$ such that:*

(1) $\det(\mathbf{M})\neq 0$*;*

(2) all the $k\times k$ minors are non-zero ($0<k<n$);

(3) all sextuple (x,y,z,a,b,c) such that

- $0<a,\ b,\ c\leq n$*;*
- $a+b+c=2n$*;*
- $a\geq b\geq c$*;*
- $0\leq x,y,z\leq n$*;*
- $x+y+z=n$*;*
- $x<a,\ y<b,\ z<c$*;*

fit $\mathbf{M}$.

Then any 4-related vectors are totally related if and only if they are coupled.

Thanks to Theorem 6.6 and Remark 6.2, we have the following

Corollary 6.7. *In the hypothesis of Theorem 6.6, any map is* 4*-extendible.*

6.2 Other embeddings of this kind

We can also build other embeddings similar to those described in previous sections. The main objective in these constructions is to identify the right compromise between computational feasibility and quantity of information that can be obtained. In Section 3.2, we constructed the embedding ε that has been useful to make linear the S-box maps which are the classical non-linear maps of a cryptosystem. We had to abandon the linearity of MixColumns (for AES) and the pLayer (in case of PRESENT). In order to use some more information about the MixColumns (or the pLayer for PRESENT), we have considered the embedding given in Section 4:

$$\alpha(v) = (\varepsilon(v), \varepsilon(\mathbf{M}v), \dots, \varepsilon(\mathbf{M}^{t-1}v)),$$

where $\mathbf{M}$ is the full Mixing Layer. The strength of this enbedding is that we can exploit the low order of $\mathbf{M}$ to force the linearity of $\mathbf{M}$. The disadvantages are that we have lost some computational efficiency and that the S-box is non-linear again (but with a lower non-linearity).

For AES, we considered also the embedding given by

$$\alpha(v) = (\varepsilon(v), \varepsilon(\mathsf{MC}(v)), \dots, \varepsilon(\mathsf{MC}^3(v))),$$

since the order of the MixColumns is equal to 4 and the MixColumns operation was the only to be non-linear in Section 3.2. Unfortunately, in this context both the ShiftRows and the parallel maps are non-linear and so we put aside this idea.

Although the following two embeddings could provide a lot of information about a cryptosystem,

- $\alpha(v) = (\varepsilon(v), \varepsilon((\mathbf{M} \circ \mathrm{Sbox})v), \dots, \varepsilon((\mathbf{M} \circ \mathrm{Sbox})^{t-1}v)) \quad (t = \mathsf{o}(\mathbf{M} \circ \mathrm{Sbox}))$
- $\alpha(v) = (\varepsilon(v), \varepsilon((\gamma\lambda\sigma_k)v), \dots, \varepsilon((\gamma\lambda\sigma_k))^{t-1}v)) \quad (t = \mathsf{o}(\gamma\lambda\sigma_k))$

they are very impractical, since the order of $(\mathbf{M} \circ \mathrm{Sbox})$ and of $(\gamma\lambda\sigma_k)$ is huge.

6.3 On complete linearizations of AES

Let $\mathcal{C}$ be any block cipher such that the plain-text space $\mathcal{M}$ coincides with the cipher space. Let $\mathcal{K}$ be the key space. Any key $k \in \mathcal{K}$ induces a permutation τ_k on $\mathcal{M}$. Since $\mathcal{M}$ is usually $V = (\mathbb{F}_2)^n$ for some $n \in \mathbf{N}$, we can consider $\tau_k \in \mathrm{Sym}(V)$. We denote by $\Gamma = \Gamma(\mathcal{C})$ the subgroup of $\mathrm{Sym}(V)$ generated by all the τ_k's. Unfortunately, the knowledge of $\Gamma(\mathcal{C})$ is out of reach for the most important block ciphers, such as the AES [Nat01] and the DES [Nat77]. However, researchers have been able to compute another related group. Suppose

that $\mathcal{C}$ is the composition of l rounds (the division into rounds is provided in the document describing the cipher). Then any key k would induce l permutations, $\tau_{k,1}, \ldots, \tau_{k,l}$, whose composition is τ_k. For any round h, we can consider $\Gamma_h(\mathcal{C})$ as the subgroup of $\mathrm{Sym}(V)$ generated by the $\tau_{k,h}$'s (with k varying in $\mathcal{K}$). We can thus define the group $\Gamma_\infty = \Gamma_\infty(\mathcal{C})$ as the subgroup of $\mathrm{Sym}(V)$ generated by all the Γ_h's. Obviously, $\Gamma \leq \Gamma_\infty$. Group Γ_∞ is traditionally called the *group generated by the round functions* with independent sub-keys. This group is known for some important ciphers, for example we have

Proposition 6.8 ([SW08],[Wer02])**.**

$$\Gamma_\infty(\mathrm{AES}) = \mathrm{Alt}((\mathbb{F}_2)^{128}).$$

It is very likely (and it is common belief among researchers) that $\Gamma_{AES} = \Gamma_\infty(\mathrm{AES}) = \mathrm{Alt}((\mathbb{F}_2)^{128})$. Assuming this, we discuss in this section the possibility of viewing Γ_{AES} as a subgroup of $\mathrm{GL}(V)$ with V of small dimension. In Cryptography it is customary to present estimates as powers of two, so our problem becomes to find the smallest ℓ such that Γ_{AES} can be linearized in $\mathrm{GL}((\mathbb{F}_2)^{2^\ell})$. A classical proof is given in [Wag76] that $\ell = 128$. We feel desirable to obtain a result with a simpler proof. Our estimate is weaker than Wagner's, but strong enough to show the linearization infeasibility.

There are two obvious ways to show that a finite group A cannot be contained (as isomorphic image) in a finite group B. The first is to show that $|A| > |B|$, the second is to show that there is $\eta \in A$ such that its order is strictly larger than the maximum element order in B. Subsection 6.3.1 presents our result using the first approach and we show that $\ell \geq 67$, which is more than enough to ensure the infeasibility of the linearization attack. This subsection's argument is completely elementary. Subsection 6.3.2 present our result using the second approach and we show again that $\ell \geq 67$. It is interesting that, although here some more advanced argument is needed (results in number theory), we reach the same estimate.

6.3.1 First approach

In this subsection we show that the order of $\mathrm{Alt}((\mathbb{F}_2)^{128})$ is strictly larger than the order of $\mathrm{GL}(V)$, with $V = (\mathbb{F}_2)^{2^{66}}$, so that $\ell \geq 67$.

We begin with showing a lemma.

Lemma 6.9. *The following inequality holds*

$$2^{(2^7)^{19}} < 2^{128}! < 2^{(2^7)^{20}}.$$

Proof. Let $n = 2^7$, we have to show $2^{n^{19}} < 2^n! < 2^{n^{20}}$. We first show that $2^{n^{19}} < 2^n!$. The following inequality holds for $1 \leq i \leq n-2$ and $1 \leq h \leq 2^{n-i}$

$$\frac{1}{2^{n-i}} \geq \frac{1}{2^{n-i+1} - h}\,. \tag{14}$$

Clearly

$$\begin{aligned} 2^n! > 2^{n^{19}} &\iff 2^n(2^n-1)! > 2^n \cdot 2^{n^{19}-n} \\ &\iff (2^n-1)(2^n-2)! > 2^{n^{19}-n} \cdot \frac{2^n-1}{2^n-1} \quad . \end{aligned}$$

We apply (14) with $i = 1$ and $h = 1$ and so we must prove

$$(2^n-1)(2^n-2)! > 2^{n^{19}-n} \cdot \frac{2^n-1}{2^{n-1}},$$

i.e. $(2^n-2)! > 2^{n^{19}-n-(n-1)}$. We use the same inequality for all $2 \leq h \leq 2^{n-1}$ and we obtain that we must verify $(2^{n-1}-1)! > 2^{n^{19}-n-2^{n-1}(n-1)}$. Then we proceed by applying (14) for all $2 \leq i \leq n-2$ and all $1 \leq h \leq 2^{n-i}$, so that we need only to prove

$$(2^{n-(n-1)}-1)! \geq 2^{n^{19}-n-\sum_{i=1}^{n-1} 2^{n-i}(n-i)}.$$

In other words, we have to prove

$$1 > 2^{n^{19}-n-\sum_{i=1}^{n-1} 2^{n-i}(n-i)}, \quad \text{that is,} \quad 0 > n^{19} - n - \sum_{i=1}^{n-1} 2^{n-i}(n-i). \quad (15)$$

But a direct check shows that the right-hand size of (15) holds when $n = 2^7$.

We are left to demonstrate the following inequality: $2^n! < 2^{n^{20}}$.
We proceed by induction for $2 \leq n \leq 2^7$. In this range a computer computation shows that

$$n^{20} + 2^n n + 2^n < (n+1)^{20}. \quad (16)$$

When $n = 2$, we have $2^2! < 2^{2^{20}}$. Suppose that $2^n! < 2^{n^{20}}$ and $n \leq 2^7$. We have to prove that $2^{(n+1)}! < 2^{(n+1)^{20}}$. Since $2^{n+1}! = (2^n \cdot 2)! = 2^n!(2^n + 1) \cdots (2^n + 2^n)$, we have

$$2^n!(2^n+1)\cdots(2^n+2^n) < 2^{n^{20}+n+1} \cdot (2^n+2)\cdots(2^n+2^n) \leq 2^{n^{20}+2^n(n+1)} = 2^{n^{20}+2^n n+2^n}$$

and, applying (16), we get $2^{n^{20}+2^n n+2^n} < 2^{(n+1)^{20}}$.
Then the claimed inequality $2^{n+1}! < 2^{(n+1)^{20}}$ follows. □

Our result is contained in the following proposition.

Proposition 6.10. *Let $V = (\mathbb{F}_2)^{2^\ell}$ with $\ell \geq 2$. If $G < \mathrm{GL}(V)$, with G isomorphic to $\mathrm{Alt}((\mathbb{F}_2)^{128})$, then $\ell \geq 67$.*

Proof. If $G < \mathrm{GL}(V)$, then $|G| \leq |\mathrm{GL}(V)|$. But $|\mathrm{Sym}((\mathbb{F}_2)^{128})| = 2^{128}! > 2^{2^{133}}$ thanks to Lemma 6.9 and so

$$|G| = |\mathrm{Alt}((\mathbb{F}_2)^{128})| = \frac{|\mathrm{Sym}((\mathbb{F}_2)^{128})|}{2} > \frac{2^{2^{133}}}{2} = 2^{2^{133}-1} > 2^{2^{132}} > |\mathrm{GL}((\mathbb{F}_2)^{2^{66}})|\,.$$

Therefore, $\ell = 66$ is not large enough. □

Remark 6.11. We could improve the previous bound to $\ell \geq 68$ by using the finite version of the Stirling fomula:

$$n\log_2 n - n\log_2(e) \leq \log_2(n!) \leq n\log_2 n - n\log_2(e) + \log_2 n, \quad \left(\frac{n}{e}\right)^n \leq n! \leq n\left(\frac{n}{e}\right)^n.$$

6.3.2 Using the order of the elements

In this subsection we compare the maximum order of elements in the two groups $\mathrm{Alt}((\mathbb{F}_2)^{128})$ and $\mathrm{GL}((\mathbb{F}_2)^{2^\ell})$. We use permutations of even order. We denote by $\mathsf{o}(\sigma)$ the order of any permutation σ, with $\sigma \in \mathrm{Alt}((\mathbb{F}_2)^{128})$ or $\sigma \in \mathrm{GL}((\mathbb{F}_2)^{2^\ell})$.

The best available result for $\mathrm{GL}((\mathbb{F}_2)^{2^\ell})$ is given by the following theorem

Theorem 6.12 ([Dar08])**.** *Let $\sigma \in \mathrm{GL}((\mathbb{F}_2)^N)$, with $\mathsf{o}(\sigma)$ is even and $N \geq 4$. Then*

$$\mathsf{o}(\sigma) \leq 2(2^{N-2} - 1) = 2^{N-1} - 2.$$

Moreover, there is $\sigma \in \mathrm{GL}((\mathbb{F}_2)^N)$ whose order attains the upper bound.

Proof. It comes directly from Theorem 1 in [Dar08], with $p = q = 2$ and $N \geq 4$ (so point (a) and (b) do not apply). □

As regards the order of the elements in $\mathrm{Alt}((\mathbb{F}_2)^{128})$, we would like to use the following theorem

Theorem 6.13 ([DM96])**.** *Let $\nu \geq 3$ and $n = 2^\nu$. Then $\mathrm{Alt}((\mathbb{F}_2)^\nu)$ contains an element η of order (strictly) greater then $e^{\sqrt{(1/4)n\ln n}}$.*

The previous theorem is the special case of Theorem 5.1.A at p.145 in [DM96] when $q = 2$.

In order to be able to compare the two estimates coming from Theorem 6.12 and Theorem 6.13, we rewrite Theorem 6.13 as follows, in order to have $\mathsf{o}(\eta)$ even. Our proof is an easy adaption of the proof contained in [DM96].

Theorem 6.14. *Let $\nu \geq 7$ and $n = 2^\nu$. Then $\mathrm{Alt}((\mathbb{F}_2)^\nu)$ contains an element η with $\mathsf{o}(\eta) > e^{\sqrt{(1/4)n\ln n}}$ and $\mathsf{o}(\eta)$ even.*

Proof. Let z be a prime number such that $4 + \sum_{3\leq p\leq z} p \leq n$, where the sum runs over (distinct odd) prime numbers. Then $\mathrm{Alt}((\mathbb{F}_2)^\nu)$ contains an element $\eta_z = \sigma\sigma'\sigma_3\cdots\sigma_p\cdots\sigma_z$ such that: σ and σ' are transpositions, σ_p is a cycle of length p, and all cycles $\{\sigma, \sigma', \sigma_3, \ldots, \sigma_z\}$ act on disjoint subsets of $(\mathbb{F}_2)^\nu$.

In other words, the non-trivial cycles of η_z are two transpositions and some cycles with length $3, \ldots, z$. As a consequence, the order of η_z is $2\prod_{3\leq p\leq z} p$.

We are going to show that there is $z \in \mathbb{N}$ such that

$$4 + \sum_{2<p\leq z} p \leq n \qquad \text{and} \qquad (\vartheta(z))^2 > \frac{1}{4} n \ln(n)$$

where $\vartheta(z) = \ln(\mathsf{o}(\eta_z)) = \ln(2) + \sum_{2<p\leq z} \ln(p)$; in the following we consider $\vartheta^*(z) = \vartheta(z) - \ln(2) = \sum_{2<p\leq z} \ln(p)$.

Since $n \geq 2^7$, we note that $4 + \sum_{2<p\leq 19} p = 79 < 128 \leq n$.

Let $f(z) = \frac{z}{\ln(z)}$. Since $f(z)$ is an increasing function for real $z > e$, in case z is real and $z \geq 19$, we have that

$$f(4)\ln(4) + f(3)\ln(3) = 7 < f(19)\ln(3) \leq f(z)\ln(3) \tag{17}$$

and so we can write (if $z \geq 19$ and $z \in \mathbb{R}$)

$$\begin{aligned} 4 + \sum_{2<p\leq z} p &= f(4)\ln(4) + \sum_{2<p\leq z} f(p)\ln(p) \\ &= f(4)\ln(4) + f(3)\ln(3) + \sum_{3<p\leq z} f(p)\ln(p) \\ &< f(z)\ln(3) + \sum_{3<p\leq z} f(z)\ln(p) \\ &= \sum_{2<p\leq z} f(z)\ln(p) = f(z) \sum_{2<p\leq z} \ln(p) = f(z)\vartheta^*(z). \end{aligned}$$

We shall choose $\bar{z} \geq 19$ such that $f(\bar{z})\vartheta^*(\bar{z}) = n$. Such a $\bar{z}$ exists because $f(19)\vartheta^*(19) < 100 < n$ and $f(z)\vartheta^*(z)$ is an increasing function assuming all values.
Since $\vartheta^*(z) > z/2$ for all $z \geq 19$, we have

$$n = \frac{\bar{z}\vartheta^*(\bar{z})}{\ln(\bar{z})} < \frac{2(\vartheta^*(\bar{z}))^2}{\ln(2\vartheta^*(\bar{z}))} = \frac{4(\vartheta^*(\bar{z}))^2}{2\ln(2\vartheta^*(\bar{z}))} = f(4(\vartheta^*(\bar{z}))^2)\,.$$

However we also have $f(n\ln(n)) < n$. Since f is an increasing function, this shows that $n\ln(n) < 4(\vartheta^*(\bar{z}))^2 < 4(\vartheta(\bar{z}))^2$. It is now enough to consider $\tilde{z}$ as the largest prime smaller than $\bar{z}$. □

Now, we compare the estimate from Theorem 6.12 and Theorem 6.13. Take $n = 2^{128}$ and $\eta \in \mathrm{Alt}((\mathbb{F}_2)^{128})$ such that $\mathsf{o}(\eta) \geq e^t$ ($\mathsf{o}(\eta)$ even), where $t = \sqrt{(1/4)n\ln n} = \sqrt{(1/4)2^{128}\ln(2^{128})}$.
Since

$$e^t = e^{\sqrt{2^{126}128\ln 2}} = e^{\sqrt{2^{133}\ln 2}} = (e^{\sqrt{2\ln 2}})^{2^{66}},$$

by replacing e with $2^{\log_2 e}$, we obtain

$$e^t = (2^{(\log_2 e)\sqrt{2\ln 2}})^{2^{66}} = 2^{2^{66}(\log_2 e)\sqrt{2\ln 2}} = 2^{2^{66}\varepsilon},$$

where $\varepsilon \in \mathbb{R}$ is circa 1.69. According to Theorem 6.14, the order of η is at least $\mathtt{o}(\eta) \geq e^{2^{66}\varepsilon}$. If $\mathrm{Alt}((\mathbb{F}_2)^{128}) \subset \mathrm{GL}((\mathbb{F}_2)^N)$, we then need the the smallest N such that $\mathtt{o}(\eta) \leq (2^{N-1}-2)$ (Theorem 6.12). In other words we have to see when the following inequality holds

$$\mathtt{o}(\eta) = 2^{2^{66}\varepsilon} \leq 2^{N-1} - 2. \tag{18}$$

We observe that

- if $N = 2^{66}$, then (18) is false, since $2^{2^{66}\varepsilon} > 2^{2^{66}} > 2^{2^{66}-1} - 2$;
- if $N = 2^{67}$, then (18) is true, since $2^{2^{66}\varepsilon} < 2^{2^{66}(1.7)} < 2^{2^{67}-1} - 2$.

Therefore, we need at least $\ell \geq 67$ in order to embed $\mathrm{Alt}((\mathbb{F}_2)^{128}) \subset \mathrm{GL}(V)$, which is exactly the same value as in Proposition 6.10.

Remark 6.15. It is shown in Landau [Lan03] that the maximum order of an element in $\mathrm{Sym}((\mathbb{F}_2)^\nu)$ is asymptotic to $e^{\sqrt{n \ln n}}$ as $n \to \infty$ (with $n = 2^\nu$). Assuming this, we observe that we could slightly improve the value of ℓ we need to $\ell \geq 68$, which is the same as Remark 6.11.

Acknowledgments

A large part of these results comes from the first author's Ph.D thesis, after some initial insights by the third author. The first author would like to thank the second author (her supervisor).

These results have been presented in a few talks (2007: Trento; 2008: Cork, Pisa; 2009: Trento; 2010: Marseille, Torino) and several scientific discussions with colleagues. The authors would like to thank the following people for their valuable comments and suggestions: G. Bertoni, A. Caranti, F. Dalla Volta, O. Dunkelman, P. Fitzpatrick, P. Fragneto, P. Gianni, L. Maines, T. Mora, L. Perret, C. Traverso, R. Wernsdorf.

For their help in the attack implementation the authors thank E. Bertolazzi and F. Caruso.

The initial discussion about this work has been supported by the STMicroelectronics contract " Complexity issues in algebraic Coding Theory and Cryptography". Further discussion took place during the Special Semester on Groebner Bases (2006), organized by RICAM, *Austrian Academy of Sciences* and RISC, Linz, Austria.

Part of this research has been funded by: `Provincia Autonoma di Trento grant` *"PAT-CRS grant"*, `MIUR grant` *"Algebra Commutativa, Combinatoria e Computazionale"*, `MIUR grant` *"Rientro dei Cervelli"*.

References

[ABK98] R. J. Anderson, E. Biham, and L.R. Knudsen, *Serpent: A new block cipher proposal*, Proc. of FSE 1998, LNCS, vol. 1372, Springer, 1998,

pp. 222–238.

[AKL+07] A. Andrey Bogdanov, L. R. Knudsen, G. Leander, C. Paar, A. Poschmann, M. J. B. Robshaw, Y. Seurin, and C. Vikkelsoe, *PRESENT: An ultra-lightweight block cipher*, Proc. of CHES 2007, LNCS, vol. 4727, Springer, 2007, pp. 450–466.

[BB02] E. Barkan and E. Biham, *In how many ways can you write Rijndael?*, Proc. of ASIACRYPT 2002, LNCS, vol. 2501, 2002, pp. 160–175.

[BDK05] E. Biham, O. Dunkelman, and N. Keller, *Related-key boomerang and rectangle attacks*, Proc. of EUROCRYPT 2005, LNCS, vol. 3494, 2005, pp. 507–525.

[BDK+10] A. Biryukov, O. Dunkelman, N. Keller, D. Khovratovich, and A. Shamir, *Key Recovery Attacks of Practical Complexity on AES-256 Variants with up to* 10 *rounds*, Proc. of EUROCRYPT 2010, LNCS, 2010.

[Bir04] A. Biryukov, *The Boomerang Attack on 5 and 6 round Reduced AES*, Proc. of AES4, 2004.

[BK00] E. Biham and N. Keller, *Cryptanalysis of reduced variants of Rijndael*, Proc. of AES3, 2000.

[BK09] A. Biryukov and D. Khovratovich, *Related-key Cryptanalysis of the Full AES-192 and AES-256*, Tech. report, IACR, 2009, http://eprint.iacr.org/2009/317.

[CDS09] A. Caranti, F. Dalla Volta, and M. Sala, *On some block ciphers and imprimitive groups*, AAECC **20** (2009), no. 5-6, 229–350.

[CKK+01] J.H. Cheon, M. Kim, K. Kim, J.Y. Lee, and S. Kang, *Improved impossible differential cryptanalysis of Rijndael and Crypton*, Proc. of ICISC 2001, LNCS, vol. 2288, 2001, pp. 39–49.

[CMR07] C. Cid, S. Murphy, and M. J. B. Robshaw, *Algebraic aspects of the Advanced Encryption Standard*, Springer, 2007.

[CW09] C. Cid and R. P. Weinmann, *Block ciphers: algebraic cryptanalysis and Gröbner bases*, Gröbner Bases, Coding, and Cryptography (M. Sala, T. Mora, L. Perret, S. Sakata, and C. Traverso, eds.), RISC Book Series, Springer, Heidelberg, 2009, p. to appear.

[CYK09] D. L. Cook, M. Yung, and A. D. Keromytis, *Elastic block ciphers: method, security and instantiations*, Int. J. Inf. Sec **8** (2009), no. 3, 211–231.

[Dar08] M. R. Darafsheh, *The maximum element order in the groups related to the linear groups which is a multiple of the defining characteristic*, Finite Fields Appl. **14** (2008), no. 4, 992–1001.

[DM96] J. D. Dixon and B. Mortimer, *Permutation groups*, vol. 163, Springer-Verlag, 1996.

[DR98] J. Daemen and V. Rijmen, *AES proposal: Rijndael*, Tech. report, NIST, 1998.

[DR02] ———, *The Design of Rijndael*, Springer, 2002.

[FKL+00] N. Ferguson, J. Kesley, S. Lucks, B. Schneier, M. Stay, D. Wagner, and D. Whitinf, *Improved cryptanalysis of Rijndael*, Proc. of FSE 2000, LNCS, vol. 1978, Springer, 2000, pp. 213–230.

[GM00] H. Gilbert and M. Minier, *A collision attack on seven rounds of Rijndael*, Proc. of AES3, 2000.

[Lan03] E. Landau, *Ueber die maximalordung der permutation gegbenen grades*, Arch. der Math. und Phys. **5** (1903), no. 3, 92–103.

[LN97] R. Lidl and H. Niederreiter, *Finite fields*, Encyclopedia of Mathematics and its Applications, Cambridge University Press, 1997.

[Mai09] Lara Maines, *Una debole rappresentazione del gruppo simmetrico*, Master's thesis (laurea specialistica), University of Trento, Department of Mathematics, 2009.

[MMM04] T. Migler, K. E. Morrison, and O. Mitchell, *Weight and rank of matrices over finite fields*, Tech. report, arxiv, 2004.

[MR02] S. Murphy and M. J. B. Robshaw, *Essential algebraic structure within the AES*, Proc. of CRYPTO 2002, LNCS, vol. 2442, Springer, 2002, pp. 1–16.

[MRS10] L. Maines, A. Rimoldi, and M. Sala, *On a weak notion of group representations*, work in progress (2010), 20.

[Nat77] National Bureau of Standards, *The Data Encryption Standard*, Federal Information Processing Standards Publication (FIPS) 46, 1977.

[Nat01] National Institute of Standards and Technology, *The Advanced Encryption Standard*, Federal Information Processing Standards Publication (FIPS) 197, 2001.

[NIS00] *A statistical test suite for random and pseudorandom number generators for cryptographic applications*, Special Publication SP 800-22, NIST, 2000.

[RSB10] A. Rimoldi, M. Sala, and E. Bertolazzi, *Do AES encryptions act randomly?*, Tech. report, arxiv, november 2010, http://arxiv.org/abs/.

[Sha49] C. E. Shannon, *Communication theory of secrecy systems*, Bell System Tech. J. **28** (1949), 656–715.

[Sot98] J. J. Soto, *Randomness testing of the AES candidate algorithms*, Proc. of AES candidate conference I (National Institute of Standards and Technology, ed.), NIST, 1998, p. 9.

[Sti95] D. R. Stinson, *Cryptography, Theory and Practice*, CRC Press, 1995.

[SW08] R. Sparr and R. Wernsdorf, *Group theoretic properties of Rijndael-like ciphers*, Discrete Appl. Math. **156** (2008), no. 16, 3139–3149.

[TZ05] I. Toli and A. Zanoni, *An algebraic interpretation of AES-128*, Proc. of AES 2004, LNCS, vol. 3373, Springer, 2005, pp. 84–97.

[Wag76] A. Wagner, *The faithful linear representation of least degree of S_n and A_n over field of characteristic* 2., Math. Z. **151** (1976), no. 2, 127–137.

[Wer02] R. Wernsdorf, *The round functions of Rijndael generate the alternating group*, Fast software encryption, Lect. Notes Comput. Sci., vol. 2365, Springer, Berlin, 2002, pp. 143–148.